

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



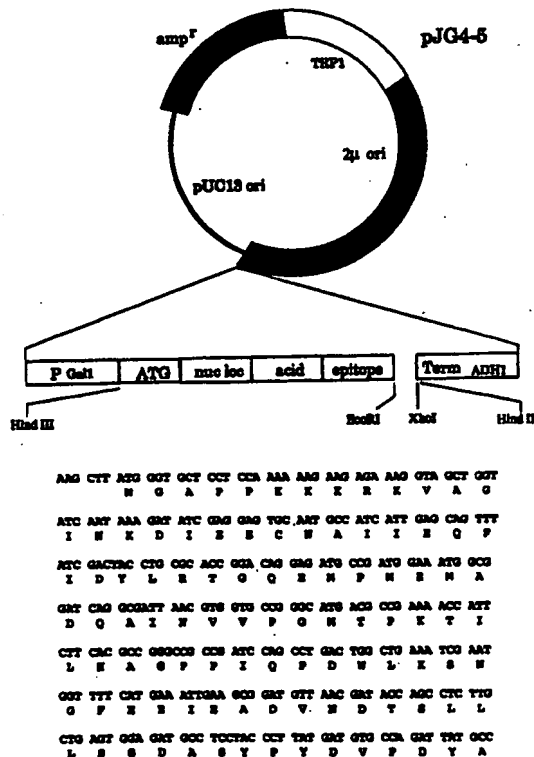
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N	A2	(11) International Publication Number: WO 95/33819 (43) International Publication Date: 14 December 1995 (14.12.95)
(21) International Application Number: PCT/US95/07113 (22) International Filing Date: 2 June 1995 (02.06.95) (30) Priority Data: 08/253,155 2 June 1994 (02.06.94) US (71) Applicant: MITOTIX, INC. [US/US]; Building 600, One Kendall Square, Cambridge, MA 02139 (US). (72) Inventors: DRAETTA, Giulio; 27 Arlington Street, Winchester, MA 01890 (US). GYURIS, Jeno; 11 Cardinal Street, Winchester, MA 01890 (US). (74) Agents: VINCENT, Matthew, P. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).		(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>

(54) Title: **CDK4 BINDING PROTEINS**

(57) Abstract

The present invention relates to the discovery of novel proteins of mammalian origin which can associate with the human cyclin dependent kinase 4 (CDK4).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

CDK4 Binding Proteins

Background of the Invention

Passage of a mammalian cell through the cell cycle is regulated at a number of key control points. Among these are the points of entry into and exit from quiescence (G_0), the restriction point, the G_1/S transition, and the G_2/M transition (for review, see Draetta (1990) *Trends Biol Sci* 15:378-383; and Sherr (1993) *Cell* 73:1059-1065). For a cell to pass through a control point and enter the next phase of the cell cycle, it must complete all of the events of the preceding cell cycle phase and, in addition, satisfy a number of check-point controls. Such controls act, for example, to ensure that DNA replication has been successfully completed before the onset of mitosis. Ultimately, information from these check-point controls is integrated through the regulated activity of a group of related kinases, the cyclin-dependent kinases (CDKs). Once a phase of the cell cycle has been successfully completed, phosphorylation of a critical substrates by activated CDKs allow passage of a cell cycle transition point and execution of the next cell cycle phase.

The ordered activation of the different CDKs constitutes the basic machinery of the cell cycle. The activity of CDKs is controlled by several mechanisms that include stimulatory and inhibitory phosphorylation events, and complex formation with other proteins. To become active, CDKs require the association of a group of positive regulatory subunits known as cyclins (see, for example, Nigg (1993) *Trends Cell Biol.* 3:296). In particular, human CDK4 exclusively associates with the D-type cyclins (D1, D2, and D3) (Xiong et al. (1992) *Cell* 71:505; Xiong et al. (1993) *Genes and Development* 7:1572; and Matsushime et al. (1991) *Cell* 65:701) and, conversely, the predominant catalytic partner of the D-type cyclins is the CDK4 kinase (Xiong et al. (1992) *Cell*). The complexes formed by CDK4 and the D-type cyclins have been strongly implicated in the control of cell proliferation during the G1 phase (Motokura et al. (1993) *Biochem. Biophys. Acta.* 1155:63-78; Sherr (1993) *Cell* 73:1059-1065; Matsushimi et al. (1992) *Cell* 71:323-334; and Kamb et al. (1994) *Science* 264:436-440).

Summary of the Invention

The present invention relates to the discovery of novel proteins of mammalian origin which can associate with the human cyclin dependent kinase 4 (CDK4). As described herein, a CDK4-dependent interaction trap assay was used to isolate a number of proteins which bind CDK4, and which are collectively referred to herein as "CDK4-binding proteins" or "CDK4-

5 BPs". In particular embodiments of the present invention, human genes have been cloned for an apparent kinase (clone #225), an apparent isopeptidase (clone #269), an apparent protease (clone #71), a human cdc37 (clone # 269), a selectin-like protein (clone #11). The present invention, therefore, makes available novel proteins (both recombinant and purified forms), recombinant genes, antibodies to the subject CDK4-binding proteins, and other novel reagents and assays for diagnostic and therapeutic use.

10 One aspect of the invention features a substantially pure preparation of a CDK4-binding protein, or a fragment thereof. In preferred embodiments: the protein comprises an amino acid sequence at least 70% homologous to the amino acid sequence represented by one of SEQ ID Nos. 25-48; the polypeptide comprises an amino acid sequence at least 80% homologous to the amino acid sequence represented by one of SEQ ID Nos. 25-48; the polypeptide comprises an amino acid sequence at least 90% homologous to the amino acid sequence of one of SEQ ID Nos. 25-48; the polypeptide comprises an amino acid sequence identical to the amino acid sequence of one of SEQ ID Nos. 25-48. In a preferred
15 embodiment: the fragment comprises at least 5 contiguous amino acid residues of one of SEQ ID Nos. 25-48; the fragment comprises at least 20 contiguous amino acid residues of one of SEQ ID Nos. 25-48; the fragment comprises at least 50 contiguous amino acid residues of one of SEQ ID Nos. 25-48. In a preferred embodiment, the fragment comprises at least a portion of the CDK4-BP which binds to a CDK, e.g. CDK4, e.g. CDK6, e.g. CDK5.

20 Yet another aspect of the present invention concerns an immunogen comprising the CDK4-binding protein, or a fragment thereof, in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for the subject CDK4-BP; e.g. a humoral response, eg. an antibody response; e.g. a cellular response.

25 A still further aspect of the present invention features an antibody preparation specifically reactive with an epitope of the CDK4-BP immunogen.

30 Another aspect of the present invention features a recombinant CDK4-binding protein, or a fragment thereof, comprising an amino acid sequence which is preferably: at least 70% homologous to one of SEQ ID Nos. 25-48; at least 80% homologous to one of SEQ ID No. 25-48; at least 90% homologous to one of SEQ ID No. 25-48. In a preferred embodiment, the recombinant CDK4-BP functions in one of either role of an agonist of cell cycle regulation or an antagonist of cell cycle regulation.

35 In one embodiment, the subject CDK4-BP is a protease. In preferred embodiments: the protease mediates degradation of cellular proteins, e.g. cell-cycle regulatory proteins, e.g. CDK4-associated proteins, e.g. cyclins, e.g. D-type cyclins; the protease affects the cellular half-life of a cell-cycle regulatory protein, e.g. a CDK-associated protein, e.g. a cyclin, e.g. a D-type cyclin, e.g. in normal cells, e.g. in cancerous cells.

In another embodiment, the subject CDK4-BP is a kinase, e.g., a stress-activated protein kinase.

In another embodiment, the subject CDK4-BP is a *Tre* oncoprotein, e.g. an isopeptidase, e.g. a deubiquitinating enzyme.

5 In yet another embodiment, the CDK4-binding protein is a human homolog of the yeast *cdc37* gene., e.g. a protein which functions to control cell-cycle progression by integrating extracellular stimulus into cell-cycle control.

10 In a still further embodiment, the CDK4-binding protein is an adhesion molecule, e.g. related to a selectin, e.g. which is responsible for integrating information from surrounding cell-cell contacts into a checkpoint control.

In yet other preferred embodiments, the recombinant CDK4-binding protein is a fusion protein further comprising a second polypeptide portion having an amino acid sequence from a protein unrelated the CDK4-binding protein. Such fusion proteins can be functional in an interaction trap assay.

15 Another aspect of the present invention provides a substantially pure nucleic acid comprising a nucleotide sequence which encodes a CDK4-binding protein, or a fragment thereof, including an amino acid sequence at least 70% homologous to one of SEQ ID Nos. 25-48. In a more preferred embodiment, the nucleic acid encodes a protein comprising an amino acid sequence at least 70% homologous to one of SEQ ID Nos. 25-28; and more
20 preferably at least 80% homologous to one of SEQ ID No. 25-28.

In yet a further preferred embodiment, the nucleic acid which encodes a CDK4-binding protein of the present invention, or a fragment thereof, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of SEQ ID Nos. 1-24 and 49-66; more preferably to at least 20 consecutive nucleotides of said
25 SEQ ID listings; more preferably to at least 40 consecutive nucleotides of said SEQ ID listings. In a preferred embodiment, the nucleic acid which encodes a CDK4-binding protein of the present invention is provided by ATCC deposit 75788.

Furthermore, in certain preferred embodiments, nucleic acids encoding one of the subject CDK4-binding protein may comprise a transcriptional regulatory sequence, e.g. at
30 least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the CDK4-BP gene sequence so as to render the gene sequence suitable for use as an expression vector. In one embodiment, the CDK4-BP gene is provided as a sense construct. In another embodiment, the CDK4-BP gene is provided as an anti-sense construct.

The present invention also features transgenic non-human animals, e.g. mice, rabbits and pigs, which either express a heterologous CDK4-BP gene, e.g. derived from humans, or which mis-express their own homolog of a CDK4-BP gene, e.g. expression of the mouse homolog of the clone #71 protease is disrupted, e.g. expression of the mouse homolog of the clone #116 isopeptidase is disrupted, e.g. expression of the mouse homolog of the clone #225 kinase is disrupted, e.g. expression of the mouse homolog of the clone #269 cdc37 is disrupted. Such a transgenic animal can serve as an animal model for studying cellular disorders comprising mutated or mis-expressed CDK4-BP genes.

The present invention also provides a probe/primer comprising a substantially purified oligonucleotide, wherein the oligonucleotide comprises a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of one of SEQ ID Nos. 1-24 and 49-66, or naturally occurring mutants thereof. In preferred embodiments, the probe/primer further comprises a label group attached thereto and able to be detected, e.g. the label group is selected from a group consisting of radioisotopes, fluorescent compounds, enzymes, and enzyme co-factors. Such probes can be used as a part of a diagnostic test kit for identifying transformed cells, such as for measuring a level of a CDK4-BP nucleic acid in a sample of cells isolated from a patient; e.g. measuring a CDK4-BP mRNA level in a cell; e.g. determining whether a genomic CDK4-BP gene has been mutated or deleted.

Another aspect of the present invention provides a method of determining if a subject, e.g. a human patient, is at risk for a disorder characterized by unwanted cell proliferation, comprising detecting, in a tissue of the subject, the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding a CDK4-binding protein, or a homolog thereof; or (ii) the mis-expression of the CDK4-BP gene. In preferred embodiments: detecting the genetic lesion comprises ascertaining the existence of at least one of a deletion of one or more nucleotides from the gene, an addition of one or more nucleotides to the gene, an substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, a gross alteration in the level of a messenger RNA transcript of the gene, the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene, or a non-wild type level of the protein. For example, detecting the genetic lesion can comprise (i) providing a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of one of SEQ ID Nos. 1-24 and 49-66, or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g. wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the CDK4-BP

gene and, optionally, of the flanking nucleic acid sequences; e.g. wherein detecting the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR); e.g. wherein detecting the lesion comprises utilizing the probe/primer in a ligation chain reaction (LCR). In alternate embodiments, the level of the protein is detected in an immunoassay.

- 5 Other features and advantages of the invention will be apparent from the following detailed description, and from the claims. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature.
- 10 See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press:1989); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No. 4,683,195; *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames & S. J. Higgins eds. 1984); *Culture Of*
- 15 *Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Methods In Enzymology*, Vols. 154 and 155 (Wu et al. eds.), *Immunochemical Methods In*
- 20 *Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

25 ***Brief Description of the Figure***

- Figure 1 illustrates the pJG4-5 library plasmid and the invariant 107 amino acid moiety it encodes. This moiety carries (amino to carboxy termini) an ATG, an SV40 nuclear localization sequence (PPKKKRKVA), the B42 transcription activation domain, and the HA1 epitope tag (YPYDVPDYA). pJG4-5 directs the synthesis of proteins under the control of
- 30 the GAL1 promoter. It carries a 2 μ replicator and a TRP1⁺ selectable marker. Each of the CDK4 binding proteins of ATCC deposit accession number 75788 are inserted as EcoRI-XhoI fragments. Downstream of the XhoI site, pJG4-5 contains the ADH1 transcription terminator.

- Figure 2 is a table demonstrating the interaction of each of the CDK-binding proteins
- 35 with other cell cycle proteins.

Figure 3 is a table demonstrating the pattern of tissue expression for the mRNA encoding each of the subject CDK4-binding protein, as well as the message size.

Detailed Description of the Invention

5 The division cycle of eukaryotic cells is regulated by a family of protein kinases known as the cyclin-dependent kinases (CDKs). The sequential activation of individual members of this family and their consequent phosphorylation of critical substrates promotes orderly progression through the cell cycle. For example, the complexes formed by the *cyclin-dependent kinase 4* (CDK4) and the D-type cyclins have been strongly implicated in the
10 control of cell proliferation during the G1 phase, and are strong candidates for oncogenes that could be major factors in tumorigenesis. Indeed, recent evidence suggests the possibility that CDK4 may serve as a general activator of cell division in most, if not all, cells.

 The present invention, as set out below, derives from the discovery that, in addition to cyclins, p21, p16, and PCNA proteins, CDK4 is also associated with several other cellular
15 proteins (hereinafter termed "CDK4-binding proteins" or "CDK4-BPs"), which associations are important to the regulation of cell growth, cell proliferation, and/or cell differentiation.

 As described herein, a CDK4-dependent interaction trap assay was used to identify proteins that can associate with human CDK4. Surprisingly, a number of proteins were identified which interact with CDK4, and were subsequently cloned from a G₀ fibroblast
20 cDNA library. Given the central role of CDK4 early in G₁ phase, the present data suggest that CDK4 is an important multiplex receiver of signal transduction data, with multiple pathways converging on it to control various aspects of the kinases's activity, including both catalytic activity and substrate specificity. Thus, because each of the proteins identified herein act close to the point of CDK4 process control, such as by channeling converging
25 upstream signals to CDK4 or demultiplexing the activation of the CDK4 kinase activity by directing divergent downstream signal propagation from CDK4, each of the subject proteins is a potential therapeutic target for agents capable of modulating cell proliferation and/or differentiation.

 The present invention, therefore, makes available novel assays and reagents for
30 therapeutic and diagnostic uses. Moreover, drug discovery assays are provided for identifying agents which can affect the binding of one of the subject CDK-binding proteins with another cell-cycle regulatory protein, or which inhibit an enzymatic activity of the subject CDK4-binding protein. Such agents can be useful therapeutically to alter the growth and/or differentiation a cell.

To further illustrate, the clone designated #71 (Table 1 and Figure 2), corresponding to the protein represented by SEQ ID No. 31 (encoded by the nucleic acid of SEQ ID No. 7), shares certain homology with ATP-dependent proteases and is strongly suspected of possessing proteolytic activity. Accordingly, this protease may be a protease involved in degradation of cell-cycle regulatory proteins, e.g. G₁-cyclins such as cyclin D1, D2 or D3. Thus, clone 71 may be involved in regulating the cellular levels of other CDK4- or CDK6-associated proteins. For instance, the subject protease could be recruited by its interaction with CDK4 or CDK6 to a CDK4/cyclin D or CDK6/cyclin D complex in order to cause degradation of a D-type cyclin (e.g. cyclin D1). Such degradation would release the CDK for subsequent binding to another G₁ cyclin. Thus, agents which disrupt the binding of the protease to CDK4 or CDK6 can be used to prevent the proteolytic destruction of certain CDK4 or CDK6 associated cyclins, e.g. effectively increases the half-life of such cyclins. Alternatively, the present invention, by providing purified and/or recombinant forms of the protease, also facilitates identification of agents which act as mechanistic inhibitors of the protease and inhibit its proteolytic action on its substrates irrespective of its ability to bind CDK. As described in U.S. Patent Application No. 08/227,850 entitled "D1 Cyclin in G₁ Progression of Cell Growth, and Uses Related Thereto", the ability to increase the cellular level of cyclin D1, such as by inhibiting its proteolysis, can be useful in preventing unwanted cell growth in certain proliferative disorders.

In another embodiment, the CDK4-binding protein is an isopeptidase, such as a de-ubiquitinating enzyme. For instance, the clone designated #116 (Table 1 and Figure 2), corresponding to the polypeptide represented by SEQ ID No. 33 (encoded by the nucleic acid of SEQ ID No. 9) shares certain homology with previously described *Tre* oncogenes and isopeptidases, and may function as a de-ubiquitinating enzyme. As is generally understood, the activities of several cellular proteins are reversibly regulated by ubiquitination and a successive de-ubiquitination steps such that the half-life of the protein, or allosteric control of its biological function, is fine tuned by the control of the level of ubiquitination of that protein. For example, as described above, cyclin degradation by ubiquitin-mediated proteolysis is an important step in the progression of the cell cycle. Thus, the subject de-ubiquitinating enzyme may be involved in balancing the level of ubiquitinated cyclin D by antagonistically competing with ubiquitin conjugating enzymes. Thus, CDK4 may be used by the subject enzyme to provide proximity to a substrate such as cyclin D. Moreover, CDK4 may provide additional substrate proximity with other cell cycle regulatory proteins, such as those involved in regulation of Rb function. Agents which inhibit either the interaction of the de-ubiquitinating enzyme with CDK4, or which mechanistically inhibit the enzyme, can be used to disrupt the balance of ubiquitination of certain regulatory proteins.

In yet another embodiment, the CDK4-binding protein is a kinase which acts on CDK4 or other proteins which bind CDK4. For instance, the clone designated #225, corresponding to the polypeptide represented by SEQ ID No. 43 (encoded by SEQ ID No. 19) shares certain homology with other kinases of the family of stress-activated protein kinases (SAPKs) or Jun kinases (JNKs). These kinases are activated in response to a variety of cellular stresses, including treatment with tumor-necrosis factor-alpha and interleukin-beta. Thus, the subject kinase may represent a novel mechanism by which G1 phase arrest is effected in response to cellular stress. The kinase may phosphorylate either CDK4 or the bound cyclin D (other CDK4 associated protein), causing inhibition of the CDK activity and cell-cycle arrest.

In still further embodiments, the CDK4-binding protein is related to an adhesion molecule, such as a selectin. For example, the pJG4-5-CDKBP clone #11, corresponding to the partially characterized protein represented by SEQ. ID No. 25 (encoded by SEQ. ID No. 1) shares approximately 50% homology with selectin proteins, adhesion molecules which are found on epithelial and possibly lymphoid cells. Growth of normal diploid mammalian cells *in vitro*, and presumably *in vivo*, is strongly regulated by the actual cell density. Cell-cell contacts via specific plasma membrane glycoproteins has been found to be a main growth regulatory principle. Malignant growth is suggested to result from impaired function of the signal transduction pathways connected with these membrane proteins. Moreover, it has been previously noted that a major control point in fibroblast cell cycle exists at the G₀-G₁ transition and is regulated by extracellular signals including contact inhibition (Han et al. (1993) *J. Cell Biol.* 122:461-471). It is asserted here that the subject adhesion molecule is responsible for integrating information from surrounding cell contacts into a checkpoint control. Consistent with this notion, nucleic acid hybridization experiments using a probe based on SEQ. ID No. 1 have detected clone 11 mRNA in normal primary fibroblasts (e.g., WI38 and IMR90), but that clone 11 mRNA levels become undetectable in SV40 Laze T transformed fibroblasts as well as fibrocarzinom or cell lines (e.g., Hs 913T cells) - each of which have lost contact inhibition and are able to form foci. Thus, the interaction of selectin-related proteins, such as clone 11, with CDKs (e.g., CDK4, CDK5 or CDK6) is a potential therapeutic target for design of agents capable of modulating proliferation and/or differentiation. In some instances, agents which restore the function of such selectin-like proteins will be desirable to inhibit proliferation. For example, peptidomimetics based on clone 11 sequences which bind CDK4, or gene therapy vehicles which deliver the clone 11 gene, can be used to mimic the function of the wild type protein and slow progression of the cell through the G₁ phase. For instance, in addition to treatment of cancer, such agents may be used to treat hypertension, diabetic macroangiopathy or arteriosclerosis, where numerous abnormalities in vascular smooth-muscle cell (vsmc) growth is a common pathology resulting from abnormal contact inhibition and accelerated entry into the S phase.

Conversely, agents which bind clone #11 and/or other related selectins and prevent binding to a CDK can be used to prevent contact inhibition and therefore enhance proliferation (and potentially inhibit differentiation). For instance, such agents can be used to relieve contact inhibition of chondrocytes, particularly fibrochondrocytes, in order to facilitate de-differentiation of these cells into chondroblast cells which produce cartilage. Thus, therapeutic agents can be identified in assays using the subject protein which are useful in the treatment of connective tissue disorders, including cartilage repair.

In similar fashion, the CDK4-binding proteins designated as clone 61 and clone 190 are homologous to other cytoskeletal elements, such as tensin and actin-binding proteins, respectively. Recent evidence suggests that certain cytoskeletal proteins not only maintain structural integrity or provide motility for a cell, but might also be associated with signal transduction. Tensin, for example, has been implicated in signal transduction, as well as the anchor for actin filaments at the focal adhesion. Accordingly, the association of CDK4 and clones 61 and 190 can be implicated, as above, in mediating such membrane-induced events as contact inhibition, etc., such interaction being a therapeutic target for modulating, for example, cell adhesion and de-adhesion and invadopodia (e.g., invasion into the extracellular matrix) by normal and transformed cells. The interaction between these molecules and CDK4 can be one wherein CDK4 is a downstream target for apparent effector molecules. Alternatively, these proteins can be substrates for CDK complexes, the phosphorylation affecting the structure or localization of the cytoskeletal elements.

In still further embodiments, the CDK4-binding protein is a DNA binding factor involved in regulation of transcription and/or replication. For example, clones 127 and 118 (see Table 1 and Figure 2) each appear to possess zinc-finger motifs which implicate them in DNA-binding. These proteins may function as downstream targets for activation or inactivation by CDK phosphorylation, and/or to localize a CDK to DNA. Moreover, the fact that clone 127 binds strongly to p53 and Rb (Figure 2) suggests an integrated role in the G₁ checkpoint(s). In yet another embodiment, the CDK4-binding protein is an mRNA-splicing factor. For instance, clone 216 is apparently such a protein, the function of which may be modulated by the action of a CDK, or which itself may modulate the activity of a CDK.

In another embodiment, the CDK4-binding protein contains a CDK consensus phosphorylation signal, and the CDK4-BP is a CDK4 substrate and/or an inhibitor of the CDK4 kinase activity. For example, each of clones #13, #22 and #227 contain such CDK consensus sequence. Thus, these cellular proteins can be downstream substrates of CDK4 (as well as CDK6 or CDK5). Additionally, the CDK4-BP, particularly the phosphoprotein form, can serve as an inhibitor of a CDK, such as CDK4. Thus, the phosphorylated CDK4-BP could serve as a feedback loop, either from CDK4 itself or from another CDK, acting to modulate the activity of a CDK to which it binds.

In still further embodiments, the CDK4-binding protein is a human homolog of the yeast Cdc37 gene (Ferguson et al. (1988) *Nuc. Acid Res.* 14:6681-6697; and Breter et al. (1983) *Mol. Cell Biol.* 3:881-891). In particular, one embodiment of the present application is directed to the association between CDK4 and a novel human protein which we identified
5 as the mammalian homolog of the yeast gene Cdc37, (though only about 14 percent homologous) the mammalian gene being referred to herein as "*cdc37*".

Studies of the temperature-sensitive Cdc37-1 mutant in *Saccharomyces cerevisiae* suggests that Cdc37 is required for exit from G₁ phase of the cell-cycle (Reed (1980) *Genetics* 95:561-577; and Ferguson et al. (1986) *Nuc Acid Res* 14:6681-6697). Mutation or
10 deletion in yeast of the Cdc37 gene results in arrest at "START", the regulatory point in the yeast cell-cycle which in many ways resembles the G₁ restriction point and G₁/S checkpoint in mammalian cells.

While the precise function of Cdc37 in yeast is not known, our observation of the human *cdc37* binding to CDK4 and CDK6 provides an explanation for the G₁ phase arrest in
15 Cdc37-1 mutant yeast cells, and also for the role of *cdc37* in mammalian cells. It is asserted herein that the mammalian *cdc37*, and presumably the yeast Cdc37, is required for activation of cyclin-dependent kinases. The *cdc37* gene product may be required for stabilization or localization of CDKs such as CDK4, or may play a more general role in the regulation of the kinase activity, such as through allosteric regulation or a chaperon-like activity which
20 facilitates assembly of multi-protein complexes with a CDK. While not wishing to be bound by any particular theory, our results in recombinant expression systems indicate that a transient complex is formed between, for example, CDK4, cyclin D1 and *cdc37*, with *cdc37* dissociating upon phosphorylation of CDK4 by a CDK-activating kinase (CAK).

Futhermore, we have observed that the *cdc37* protein itself is apparently regulated, at
25 least in part, by phosphorylation, the phosphorylated form evidently mediating the interaction of, for example, CDK4 and cyclin D1. Using immobilized *cdc37*, several proteins which bind to *cdc37* were purified, e.g. by *cdc37* chromatography. Detecting phosphorylation of a *cdc37* substrate, a kinase activity was eluted from the *cdc37* column under a salt gradient. The active fractions were pooled, and separated by gel electrophoresis, and an in-gel kinase
30 assay was performed. Five bands, approximate molecular weights of 40kd, 42kd, 95kd, 107kd and 117kd, were identified in the gel as having kinase activity towards *cdc37*. Two of the five bands appeared as a doublet, each having a molecular weight of approximately 40 kd. This pattern has been observed previously in the literature for various members of the *erk* kinase family (for review, see Cobb et al. (1994) *Semin Cancer Biol* 5:261-8), which kinases
35 are involved in signal transduction, especially from mitogenic signals. For instance, transforming agents utilize this cascade in inducing cell proliferation. Indeed, western blot analysis revealed that these two kinase bands isolated by *cdc37* binding were the *erk-1* and

erk-2 kinases, and immunopurified forms of each of these serine/threonine kinases was found to phosphorylate (and activate) *cdc37*.

Thus, it is understood by the present invention that the human *cdc37* functions to control cell-cycle progression, perhaps by integrating extracellular stimulus into cell-cycle control, and it is therefore expected that the CDK4-*cdc37*, CDK6-*cdc37* and *erk-cdc37* interactions can be a very important target for drug design. For instance, agents which disrupt the binding of a CDK and *cdc37*, e.g., CDK4 peptidomimetic which bind *cdc37*, could be used to effect the progression of cell through G₁. Moreover, antagonistic mutants of the subject *cdc37* protein, e.g., mutants which disrupt the function of the normal *cdc37* protein, can be provided by gene therapy in order to inhibit proliferation of cells. Furthermore, the fact that the human *cdc37* homolog binds Src and p53 supports the role of *cdc37* in cell-cycle checkpoints, as well as suggesting alternate therapeutic targets, e.g., the Src-*cdc37* or p53-*cdc37* interactions.

Furthermore, it is demonstrated here for the first time that p16 is able to associate with CDK6. Previously, p16 was believed to associate exclusively with CDK4 and acted as an inhibitor of the CDK4 kinase activity. The present data strongly suggests that p16 functions in the same or similar role with respect to CDK6. Thus, the interaction between p16 and CDK6 is a potential therapeutic target for agents which (i) disrupt this interaction; (ii) mimic this interaction by binding CDK6 in a manner analogous to p16, e.g. p16 peptidomimetics which bind CDK6; or (iii) are mechanistic inhibitors of the CDK6 kinase activity. Moreover, as described below, the present invention provides differential screening assays for identifying agents which disrupt or otherwise alter the regulation of only one of either CDK4 or CDK6 without substantially affecting the other.

In general, polypeptides designated herein as CDK4-binding proteins refers to polypeptides that (i) have an amino acid sequence corresponding (identical or homologous) to all or a portion of an amino acid sequence of one of the subject CDK4-binding protein designated by SEQ ID Nos: 25-48 and (ii) which have at least one biochemical activity of that CDK4-binding protein. In preferred embodiments, a biological activity of a CDK4-binding protein can be characterized as including, in addition to those activities described above for individual clones, the ability to bind to a cyclin dependent kinase, preferably CDK4. The above notwithstanding, the biological activity of a CDK4-binding protein may be distinguished by one or more of the following attributes: an ability to regulate the cell-cycle of a eukaryotic cell, e.g. a mammalian cell cycle, e.g., a human cell cycle; an ability to regulate proliferation/cell growth of a eukaryotic cell, e.g. a mammalian cell, e.g., a human cell; an ability to regulate progression of a eukaryotic cell through G₁ phase, e.g. regulate progression of a mammalian cell from G₀ phase into G₁ phase, e.g. regulate progression of a mammalian cell through G₁ phase; an ability to regulate the kinase activity of a cyclin

dependent kinase, e.g. a CDK active in G₁ phase, e.g. CDK4, e.g. CDK6; an ability to regulate phosphorylation of an Rb or Rb-related protein by CDK4; an ability to regulate the effects of mitogenic stimulation on cell-cycle progression, e.g. regulate contact inhibition, e.g. mediate growth factor- or cytokine-induced mitogenic stimulation, e.g. regulate paracrine-responsiveness. Certain of the CDK4-binding proteins of the present invention may also have biological activities which include an ability to suppress tumor cell growth, e.g. in a tumor cell which has lost contact inhibition, e.g. in tumor cells which have paracrine feedback loops. Other biological activities of the subject CDK4-binding proteins are described herein or will be reasonably apparent to those skilled in the art. Moreover, according to the present invention, a polypeptide has biological activity if it is a specific agonist or antagonist of a naturally-occurring form of a CDK4-binding protein.

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

As used herein, the term "gene" or "recombinant gene" refers to a nucleic acid comprising an open reading frame encoding a *CDK4-binding protein* of the present invention, including both exon and (optionally) intron sequences. The term "intron" refers to a DNA sequence present in a given *cdc37* gene which is not translated into protein and is generally found between exons.

As used herein, the term "transfection" means the introduction of a nucleic acid, e.g., an expression vector, into a recipient cell by nucleic acid-mediated gene transfer. "Transformation", as used herein, refers to a process in which a cell's genotype is changed as a result of the cellular uptake of exogenous DNA or RNA, and, for example, the transformed cell expresses a recombinant form of one of the subject CDK4-binding proteins, or where anti-sense expression occurs from the transferred gene, the expression of a naturally-occurring form of the *CDK4-binding protein* is disrupted.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In general, expression vectors of utility

in recombinant DNA techniques are often in the form of "plasmids" which refer to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. In the present specification, "plasmid" and "vector" are used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in the art subsequently hereto.

"Transcriptional regulatory sequence" is a generic term used throughout the specification to refer to DNA sequences, such as initiation signals, enhancers, and promoters, which induce or control transcription of protein coding sequences with which they are operably linked. In preferred embodiments, transcription of a recombinant gene is under the control of a promoter sequence (or other transcriptional regulatory sequence) which controls the expression of the recombinant gene in a cell-type in which expression is intended. It will also be understood that the recombinant gene can be under the control of transcriptional regulatory sequences which are the same or which are different from those sequences which control transcription of the naturally-occurring form of the *CDK4-binding protein*.

As used herein, the term "tissue-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue, such as cells of a urogenital origin, e.g. renal cells, or cells of a neural origin, e.g. neuronal cells. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

As used herein, a "transgenic animal" is any animal, preferably a non-human mammal, a bird or an amphibian, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animals described herein, the transgene causes cells to express a recombinant form of a *CDK4-binding protein*, e.g. either agonistic or antagonistic forms. However, transgenic animals in which the recombinant gene is silent are also contemplated, as for example, the FLP or CRE recombinase dependent constructs described below. The "non-human animals" of the invention include vertebrates such as rodents, non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, etc. Preferred

non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians, such as members of the *Xenopus* genus, and transgenic chickens can also provide important tools for understanding, for example, embryogenesis and tissue patterning. The term "chimeric animal" is used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term "tissue-specific chimeric animal" indicates that the recombinant gene is present and/or expressed in some tissues but not others.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., a *cdc37* polypeptide or other CDK4-BP), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a *CDK4-binding protein*" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

"Homology" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny

may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding one of the subject *CDK4-binding proteins* with a second amino acid sequence defining a domain foreign to and not substantially homologous with any domain of the polypeptide making up the first sequence. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms.

The term "evolutionarily related to", with respect to nucleic acid sequences encoding each of the subject CDK4-binding proteins, refers to nucleic acid sequences which have arisen naturally in an organism, including naturally occurring mutants. The term also refers to nucleic acid sequences which, while derived from a naturally occurring gene, have been altered by mutagenesis, as for example, combinatorial mutagenesis described below, yet still encode polypeptides which have at least one activity of a CDK4-binding protein.

The term "isolated" as also used herein with respect to nucleic acids, such as DNA or RNA, refers to molecules separated from other DNAs, or RNAs, respectively, that are present in the natural source of the macromolecule. For example, isolated nucleic acids encoding the subject polypeptides preferably include no more than 10 kilobases (kb) of nucleic acid sequence which naturally immediately flanks a particular *CDK4-BP* gene in genomic DNA or mRNA, more preferably no more than 5kb of such naturally occurring flanking sequences, and most preferably less than 1.5kb of such naturally occurring flanking sequence. The term isolated as used herein also refers to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and would not be found in the natural state.

As described herein, one aspect of the invention pertains to an isolated nucleic acid having a nucleotide sequence encoding one of the subject CDK4-binding proteins, fragments thereof, and/or equivalents of such nucleic acids. The term equivalent is understood to include nucleotide sequences encoding functionally equivalent CDK4-binding proteins or functionally equivalent polypeptides which, for example, retain the ability to bind a CDK (e.g. CDK4), and which may additionally retain other activities of a CDK4-binding protein such as described herein. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will also include sequences that differ from the nucleotide sequence encoding the presently

5 claimed CDK4-binding proteins shown in any of SEQ ID Nos: 1-24 or 49-70 due to the degeneracy of the genetic code. Equivalents will also include nucleotide sequences that hybridize under stringent conditions (i.e., equivalent to about 20-27°C below the melting temperature (T_m) of the DNA duplex formed in about 1M salt) to the nucleotide sequence of a CDK4-binding protein represented by one of SEQ ID Nos: 25-48, or to a nucleotide sequence of a CDK4-BP insert of the vector pJG4-5-CDKBP (ATCC accession no. 75788). In one embodiment, equivalents will further include nucleic acid sequences derived from, and evolutionarily related to, a nucleotide sequences shown in any of SEQ ID Nos: 1-24.

10 Moreover, it will be generally appreciated that, under certain circumstances, it may be advantageous to provide homologs of the subject CDK4-binding proteins which function in a limited capacity as one of either a CDK4-BP agonists or a CDK4-BP antagonists, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects relative to treatment with agonists or antagonists which are directed to all CDK4-BP related biological activities. Such homologs of the subject CDK4-binding proteins can be generated by mutagenesis, such as by discrete point mutation(s) or by truncation. For instance, mutation can give rise to homologs which retain the substantially same, or merely a subset, of the biochemical activity of the CDK4-BP from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally occurring form of the protein. For example, homologs can be made which, relative the authentic form of the protein, competitively bind to CDK4 or other upstream or downstream binding partners of the naturally occurring CDK4-BP, but which are not themselves capable of forming productive complexes for propagating an intracellular signal or the like. When expressed in the same cell as the wild-type protein, such antagonistic mutants could be, for example, analogous to a dominant negative mutation arising in the cell. To illustrate, the homologs of the clone #71 protease might be generated to retain a protease activity, or, conversely, engineered to lack a protease activity, yet retain the ability to bind CDK4. In the instance of the latter, the catalytically inactive protease can be used to competitively inhibit the binding to CDK4 of the naturally-occurring form of the protease. In similar fashion, clone #225 homologs can be provided which, for example, are catalytically inactive as kinases, yet which still bind to a CDK. Such homolog are likely to act antagonistically to the role of the natural enzyme in cell cycle regulation, and can be used, for example, to inhibit paracrine feedback loops. Likewise, clone #116 homologs can be generated which are not capable of mediating ubiquitin levels, yet which nevertheless competitively bind CDK4 and therefore act antagonistically to the wild-type form of the isopeptidase when expressed in the same cell.

In one embodiment, the nucleic acid encodes a polypeptide which is a specific agonist (mimetic) or antagonist of a naturally occurring form of one of the subject CDK4-binding proteins. Preferred nucleic acids encode a polypeptide at least 70% homologous, more preferably 80% homologous and most preferably 85% homologous with an amino acid sequence shown in any of SEQ ID NOS: 25-48. Nucleic acids which encode polypeptides including amino acid sequences at least about 90%, more preferably at least about 95%, and most preferably identical with a sequence shown in any of SEQ ID NOS: 25-48 are also within the scope of the invention.

Certain of the nucleotide sequences shown in SEQ ID Nos. 1-24 and 49-70 encode portions of the subject CDK4-binding proteins. Therefore, in a further embodiment of the invention, the recombinant CDK4-BP genes can further include, in addition to nucleotides encoding the amino acid sequence shown in SEQ ID Nos. 25-48, additional nucleotide sequences which encode amino acids at the C-terminus and N-terminus of each protein, though not shown in those sequence listings. For instance, a recombinant CDK4-BP gene can include nucleotide sequences of a PCR fragment generated by amplifying the one of the coding sequence of one of the CDK4-BP clones of pJG4-5-CDKBP using sets of primers derived from Table 1.

Another aspect of the invention provides a nucleic acid which hybridizes under high or low stringency conditions to a nucleic acid which encodes a polypeptide having all or a portion of an amino acid sequence shown in any of SEQ ID NOS: 25-48. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

Isolated nucleic acids encoding polypeptides, as described herein, and having a sequence which differs from the nucleotide sequence shown any of SEQ ID NOS: 1-24 due to degeneracy in the genetic code are also within the scope of the invention. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC each encode histidine) may result in "silent" mutations which do not affect the amino acid sequence of the CDK4-binding protein. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject CDK4-binding proteins will exist individuals. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about

3-5% of the nucleotides) of the nucleic acids encoding a particular member of CDK4-BP family may exist among individuals of a given species due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of this invention.

5 Fragments of the nucleic acids encoding a biologically active portion of the subject CDK4-binding proteins are also within the scope of the invention. As used herein, a nucleic acid "fragment" encoding a bioactive portion of a CDK4-binding protein refers to a nucleic acid having fewer nucleotides than the nucleotide sequence encoding the entire amino acid
10 at least a portion of the biochemical function of the full-length protein, or is a specific antagonist thereof. Nucleic acid fragments within the scope of the present invention include those capable of hybridizing under high or low stringency conditions with nucleic acids from other species for use in screening protocols to detect CDK4-BP homologs, as well as those capable of hybridizing with nucleic acids from human specimens for use in detecting the
15 presence of a nucleic acid encoding one of the subject CDK4-BPs, including alternate isoforms, e.g. mRNA splicing variants. Nucleic acids within the scope of the invention may also contain linker sequences, modified restriction endonuclease sites and other sequences useful for molecular cloning, expression or purification of recombinant forms of the subject CDK4-binding proteins.

20 As indicated by the examples set out below, a nucleic acid encoding one of the subject CDK4-binding protein may be obtained from mRNA present in any of a number of eukaryotic cells. It should also be possible to obtain nucleic acids encoding the subject CDK4-binding proteins from genomic DNA obtained from both adults and embryos. For example, a gene encoding a CDK4-binding protein can be cloned from either a cDNA or a
25 genomic library in accordance with protocols herein described, as well as those generally known to persons skilled in the art. For instance, a cDNA encoding one of the subject CDK4-binding proteins can be obtained by isolating total mRNA from a cell, e.g. a mammalian cell, e.g. a human cell, including tumor cells. Double stranded cDNAs can then be prepared from the total mRNA, and subsequently inserted into a suitable plasmid or
30 bacteriophage vector using any one of a number of known techniques. A gene encoding a CDK4-binding protein can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acid of the invention can be DNA or RNA. A preferred nucleic acid is: e.g. a cDNA comprising a nucleic acid sequence represented by any one of SEQ ID Nos:
35 1-24 and 49-70; e.g. a cDNA derived from the pJG4-5-CDKBP library of ATCC deposit no. 75788.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridizes (e.g. binds) under cellular conditions, with the cellular mRNA and/or genomic DNA encoding a CDK4-binding protein so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a CDK4-binding protein. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences encoding a CDK4-binding protein. Such oligonucleotide probes are preferably modified oligonucleotide which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and is therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by van der Krol et al. (1988) *Biotechniques* 6:958-976; and Stein et al. (1988) *Cancer Res* 48:2659-2668.

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general. For such therapy, the oligomers of the invention can be formulated for a variety of modes of administration, including systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous for injection, the oligomers of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligomers may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included.

Systemic administration can also be by transmucosal or transdermal means, or the compounds can be administered orally. For transmucosal or transdermal administration,

penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligomers are formulated into conventional oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind.

This invention also provides expression vectors comprising a nucleic acid encoding one of the subject CDK4-binding proteins and operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Accordingly, the term regulatory sequence includes promoters, enhancers and other expression control elements. Exemplary regulatory sequences are described in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences that control the expression of a DNA sequence when operatively linked to it may be used in these vectors to express DNA sequences encoding the *cdc37* proteins of this invention. Such useful expression control sequences, include, for example, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage lambda, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α -mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered.

Still another aspect of the invention concerns the use of expression constructs of the subject CDK4-binding proteins in methods by which it is administered in a biologically effective carrier, e.g. any formulation or composition capable of effectively transfecting cells

in vivo with a recombinant CDK4-BP gene. Approaches include insertion of the subject gene in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors can be used to transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramicidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or CaPO_4 precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Moreover, such constructs can be used to deliver antisense expression vectors, e.g., constructs whose transcription product is complementary to at least a portion of the coding sequence of one of the subject CDK4-BP genes.

Another aspect of the present invention concerns recombinant forms of the subject CDK4-binding proteins which have at least one biological activity of a subject CDK4-binding protein, or alternatively, which are antagonists of at least one biological activity of a CDK4-BP of the present invention, including naturally occurring dysfunctional mutants. The term "recombinant protein" refers to a protein of the present invention which is produced by recombinant DNA techniques, wherein generally DNA encoding the subject CDK4-binding protein is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant gene encoding the recombinant CDK4-BP, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native CDK4-binding protein of the present invention, or an amino acid sequence similar thereto, which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring CDK4-binding protein of an organism. Recombinant proteins preferred by the present invention, comprise amino acid sequences which are at least 60% homologous, more preferably 70% homologous and most preferably 80% homologous with an amino acid sequence shown in any of SEQ ID NOS: 25-48. Polypeptides having an activity of, or which are antagonistic to, the subject CDK4-binding proteins and having at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology with a sequence of either in any of SEQ ID NOS: 25-48 are also within the scope of the invention. Thus, the present invention further pertains to recombinant forms of the subject CDK4-binding proteins which are encoded by genes derived from, e.g., a mammal, and which have amino acid sequences evolutionarily related to a subject CDK4-binding protein of any of SEQ ID NOS: 25-48, e.g., CDK4-binding proteins having amino acid sequences which have arisen naturally (e.g. by allelic variance or by differential

splicing), as well as mutational variants of *cdc37* proteins which are derived, for example, by combinatorial mutagenesis.

5 The present invention further pertains to methods of producing the subject CDK4-binding proteins. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding one of the subject CDK4-binding proteins can be cultured under appropriate conditions to allow expression of the polypeptide to occur. The polypeptide may be secreted and isolated from a mixture of host cells and medium. Alternatively, the polypeptide may be retained cytoplasmically and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts.
10 Suitable media for cell culture are well known in the art.

The recombinant CDK4-binding protein can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such polypeptide. In a preferred
15 embodiment, the recombinant CDK4-binding protein is a fusion protein containing a domain which facilitates its purification, such as a CDK4-BP-GST or poly(His)-CDK4-BP fusion protein.

Thus, a nucleotide sequence derived from the cloning of the CDK4-binding proteins of the present invention, encoding all or a selected portion of a protein, can be used to
20 produce a recombinant form of a CDK4-BP via microbial or eukaryotic cellular processes. Ligating the polynucleotide sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known intracellular proteins, e.g. p53, CDK4, RB, p16, p21, and the like. Similar procedures, or
25 modifications thereof, can be employed to prepare recombinant CDK4-binding proteins, or portions thereof, by microbial means or tissue-culture technology in accord with the subject invention.

The recombinant CDK4-BP gene can be produced by ligating a nucleic acid encoding a subject CDK4-binding protein, or a portion thereof, into a vector suitable for expression in
30 either prokaryotic cells, eukaryotic cells, or both. Expression vehicles for production of recombinant forms of the subject CDK4-binding proteins include plasmids and other vectors. For instance, suitable vectors for the expression of a CDK4-BP include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as *E. coli*. In an
35 illustrative embodiment, a CDK4-binding protein is produced recombinantly utilizing an expression vector generated by sub-cloning a gene encoding the protein from the pJG4-5-

CDKBP library (ATCC accession no. 75788) using, for example, primers based on 5' or 3' sequences of the particular pJG4-5 gene (see Table 1) and/or primers based on the flanking plasmid sequences of the pJG4-5 plasmid (e.g. SEQ ID Nos. 71 and 72).

5 A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach *et al.* (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83). These vectors can replicate in *E. coli* due the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron
10 plasmid. In addition, drug resistance markers such as ampicillin can be used.

The preferred mammalian expression vectors contain both prokaryotic sequences to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNA1/amp, pcDNA1/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived
15 vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papilloma virus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient
20 expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning: A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press:1989) Chapters 16
25 and 17. In some instances, it may be desirable to express the recombinant CDK4-binding protein by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the β -gal containing pBlueBac III).

30 When expression of a portion of one of the subject CDK4-binding proteins is desired, i.e. a truncation mutant, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from *E. coli*
35 (Ben-Bassat *et al.* (1987) *J. Bacteriol.* 169:751-757) and *Salmonella typhimurium* and its *in vitro* activity has been demonstrated on recombinant proteins (Miller *et al.* (1987) *PNAS* 84:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved

either *in vivo* by expressing CDK4-BP-derived polypeptides in a host which produces MAP (e.g., *E. coli* or CM89 or *S. cerevisiae*), or *in vitro* by use of purified MAP (e.g., procedure of Miller *et al. supra*).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. This type of expression system can be useful under conditions where it is desirable to produce an immunogenic fragment of a CDK4-binding protein. For example, the VP6 capsid protein of rotavirus can be used as an immunologic carrier protein for portions of the CDK4-BP polypeptide, either in the monomeric form or in the form of a viral particle. The nucleic acid sequence corresponding to a portion of a subject CDK4-binding protein to which antibodies are to be raised can be incorporated into a fusion gene construct which includes coding sequences for a late vaccinia virus structural protein to produce a set of recombinant viruses expressing fusion proteins comprising a portion of the protein CDK4-BP as part of the virion. It has been demonstrated with the use of immunogenic fusion proteins utilizing the Hepatitis B surface antigen fusion proteins that recombinant Hepatitis B virions can be utilized in this role as well. Similarly, chimeric constructs coding for fusion proteins containing a portion of a subject CDK4-binding protein and the poliovirus capsid protein can be created to enhance immunogenicity of the set of polypeptide antigens (see, for example, EP Publication No. 0259149; and Evans *et al.* (1989) *Nature* 339:385; Huang *et al.* (1988) *J. Virol.* 62:3855; and Schlienger *et al.* (1992) *J. Virol.* 66:2).

The Multiple Antigen Peptide system for polypeptide-based immunization can also be utilized to generate an immunogen, wherein a desired portion of a subject CDK4-binding protein is obtained directly from organo-chemical synthesis of the polypeptide onto an oligomeric branching lysine core (see, for example, Posnett *et al.* (1988) *JBC* 263:1719 and Nardelli *et al.* (1992) *J. Immunol.* 148:914). Antigenic determinants of the subject CDK4-binding proteins can also be expressed and presented by bacterial cells.

In addition to utilizing fusion proteins to enhance immunogenicity, it is widely appreciated that fusion proteins can also facilitate the expression of proteins, such as any one of the CDK4-binding proteins of the present invention. For example, a CDK4-binding protein of the present invention can be generated as a glutathione-S-transferase (GST- fusion protein). Such GST fusion proteins can enable easy purification of a CDK4-binding protein, such as by the use of glutathione-derivativized matrices (see, for example, *Current Protocols in Molecular Biology*, eds. Ausabel *et al.* (N.Y.: John Wiley & Sons, 1991)). In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site sequence at the N-terminus of the desired portion of a CDK4-binding protein, can allow purification of the poly(His)- expressed CDK4-BP-fusion protein by affinity chromatography using a Ni²⁺ metal resin. The purification leader

sequence can then be subsequently removed by treatment with enterokinase (e.g., see Hochuli *et al.* (1987) *J. Chromatography* 411:177; and Janknecht *et al.* *PNAS* 88:8972).

Techniques for making fusion genes are well known. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992).

The present invention also makes available isolated CDK4-binding proteins which are isolated from, or otherwise substantially free of other cellular or viral proteins normally associated with the protein, e.g. other cell-cycle proteins, e.g. CDKs, cyclins, p16, p21, p19 or PCNA. The term "substantially free of other cellular or viral proteins" (also referred to herein as "contaminating proteins") is defined as encompassing CDK4-BP preparations comprising less than 20% (by dry weight) contaminating protein, and preferably comprises less than 5% contaminating protein. Functional forms of the subject CDK4-binding proteins can be prepared, for the first time, as purified preparations by using, for example, a cloned gene as described herein. By "purified", it is meant, when referring to a polypeptide or DNA or RNA sequence, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins (e.g. other CDK4-BPs, or CDKs). The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (e.g., in an acrylamide gel) but not obtained either as pure (e.g. lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, e.g. acrylamide or agarose) substances or solutions. The term polypeptide, as used herein, refers to peptides, proteins, and polypeptides.

However, the subject polypeptides can also be provided in pharmaceutically acceptable carriers for formulated for a variety of modes of administration, including

systemic and topical or localized administration. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, PA. In an exemplary embodiment, the polypeptide is provided for transmucosal or transdermal delivery. For such administration, penetrants appropriate to the barrier to be permeated are used in the formulation with the polypeptide. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays or using suppositories. For topical administration, the oligomers of the invention are formulated into ointments, salves, gels, or creams as generally known in the art.

Another aspect of the invention related to polypeptides derived from the full-length CDK4-binding protein. Isolated peptidyl portions of the subject proteins can be obtained by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such polypeptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase Fmoc or t-Boc chemistry. For example, the protein can be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of, for example, CDK4 activation, such as by microinjection assays. In an illustrative embodiment, peptidyl portions of *cdc37* can be tested for CDK-binding activity or *erk*-binding, as well as inhibitory ability, by expression as, for example, thioredoxin fusion proteins, each of which contains a discrete fragment of the protein (see, for example, U.S. Patents 5,270,181 and 5,292,646; and PCT publication WO94/02502).

It is also possible to modify the structure of the subject CDK4-binding proteins for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*). Such modified polypeptides, when designed to retain at least one activity of the naturally-occurring form of the protein, are considered functional equivalents of the CDK4-binding proteins described in more detail herein. Such modified polypeptides can be produced, for instance, by amino acid substitution, deletion, or addition.

Moreover, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids can be divided into four

families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, *Biochemistry*, 2nd ed, Ed. by L. Stryer, WH Freeman and Co.:1981). Whether a change in the amino acid sequence of a polypeptide results in a functional CDK4-BP homolog can be readily determined by assessing the ability of the variant polypeptide to produce a response in cells in a fashion similar to the wild-type CDK4-BP. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of any one of the presently disclosed CDK4-binding proteins, as well as truncation mutants, and is especially useful for identifying potentially useful variant sequences which are useful in regulating cell growth or differentiation. One purpose for screening such combinatorial libraries is, for example, to isolate novel CDK4-BP homologs which function in the capacity of one of either an agonists or an antagonist of the biological activities of the wild-type ("authentic") protein, or alternatively, which possess novel activities all together. To illustrate, homologs of the clone #225 kinase can be engineered by the present method to provide catalytically inactive enzymes which maintain binding to CDK4 but which act antagonistically to the role of the native kinase in eukaryotic cells, e.g. in regulating cell growth, e.g. in regulating paracrine signal transduction. Similar embodiments are contemplated for *cdc37* polypeptides which retain the ability to bind to an *erk* kinase, e.g. *erk1* or *erk2*. Such proteins, when expressed from recombinant DNA constructs, can be used in gene therapy protocols.

Likewise, mutagenesis can give rise to CDK4-BP homologs which have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process which result in destruction of, or otherwise inactivation of, the authentic CDK4-binding protein. Such CDK4-BP homologs, and the genes which encode them, can be utilized to alter the envelope of expression for the particular recombinant CDK4 binding proteins by modulating the half-life of the recombinant protein. For instance, a short half-life can give rise to more transient biological effects associated with a particular recombinant CDK4-binding protein and, when part of an inducible expression system, can

allow tighter control of recombinant CDK4-BP levels within the cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols.

5 In a representative embodiment of this method, the amino acid sequences for a population of *cdc37* protein homologs are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, homologs from one or more species, or homologs from the same species but which differ due to mutation. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the combinatorial
10 library is produced by way of a degenerate library of genes encoding a library of polypeptides which each include at least a portion of potential *cdc37* protein sequences. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential *cdc37* nucleotide sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display).

15 There are many ways by which the library of potential homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then be ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential *cdc37*
20 sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) Tetrahedron 39:3; Itakura et al. (1981) Recombinant DNA, Proc. 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed
25 in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249:404-406; Cwirla et al. (1990) PNAS 87: 6378-6382; as well as U.S. Patents No: 5,223,409, 5,198,346, and 5,096,815).

Alternatively, other forms of mutagenesis can be utilized to generate a combinatorial
30 library. For example, CDK4-BP homologs (both agonist and antagonist forms) can be generated and isolated from a library by screening using, for example, alanine scanning mutagenesis and the like (Ruf et al. (1994) Biochemistry 33:1565-1572; Wang et al. (1994) J. Biol. Chem. 269:3095-3099; Balint et al. (1993) Gene 137:109-118; Grodberg et al. (1993) Eur. J. Biochem. 218:597-601; Nagashima et al. (1993) J. Biol. Chem. 268:2888-2892;
35 Lowman et al. (1991) Biochemistry 30:10832-10838; and Cunningham et al. (1989) Science 244:1081-1085), by linker scanning mutagenesis (Gustin et al. (1993) Virology 193:653-660; Brown et al. (1992) Mol. Cell Biol. 12:2644-2652; McKnight et al. (1982) Science 232:316);

by saturation mutagenesis (Meyers et al. (1986) Science 232:613); by PCR mutagenesis (Leung et al. (1989) Method Cell Mol Biol 1:11-19); or by random mutagenesis (Miller et al. (1992) A Short Course in Bacterial Genetics, CSHL Press, Cold Spring Harbor, NY; and Greener et al. (1994) Strategies in Mol Biol 7:32-34). Linker scanning matagenesis, particularly in a combinatorial setting, is on attractive method for identifying truncated (bioactive) forms of the protein.

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations and truncations, and, for that matter, for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of CDK4-BP homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate sequences created by combinatorial mutagenesis techniques.

In an illustrative embodiment of a screening assay, the candidate combinatorial gene products are displayed on the surface of a cell, and the ability of particular cells or viral particles to bind a CDK, such as CDK4 or CDK6, or other binding partners of that CDK4-binding protein, via this gene product is detected in a "panning assay". For instance, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell (Ladner et al., WO 88/06630; Fuchs et al. (1991) Bio/Technology 9:1370-1371; and Goward et al. (1992) TIBS 18:136-140), and the resulting fusion protein detected by panning, e.g. using a fluorescently labeled molecule which binds the CDK4-binding protein, e.g. FITC-CDK4, to score for potentially functional homologs. Cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In similar fashion, the gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical E. coli filamentous phages M13, fd, and fl

are most often used in phage display libraries, as either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) J. Biol. Chem. 267:16007-16010; Griffiths et al. (1993) EMBO J 12:725-734; Clackson et al. (1991) Nature 352:624-628; and Barbas et al. (1992) PNAS 89:4457-4461).

In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening CDK4-binding protein combinatorial libraries of the present invention. For instance, the pCANTAB 5 phagemid of the RPAS kit contains the gene which encodes the phage gIII coat protein. The combinatorial gene library can be cloned into the phagemid adjacent to the gIII signal sequence such that it will be expressed as a gIII fusion protein. After ligation, the phagemid is used to transform competent *E. coli* TG1 cells. Transformed cells are subsequently infected with M13KO7 helper phage to rescue the phagemid and its candidate gene insert. The resulting recombinant phage contain phagemid DNA encoding a specific candidate CDK4-binding protein, and display one or more copies of the corresponding fusion coat protein. The phage-displayed candidate proteins which are capable of, for example, binding CDK4, are selected or enriched by panning. For instance, the phage library can be panned on glutathione immobilized CDK4-GST fusion proteins, and unbound phage washed away from the cells. The bound phage is then isolated, and if the recombinant phage express at least one copy of the wild type gIII coat protein, they will retain their ability to infect *E. coli*. Thus, successive rounds of reinfection of *E. coli*, and panning will greatly enrich for homologs which can then be screened for further biological activities in order to differentiate agonists and antagonists.

Consequently, the invention also provides for reduction of the subject CDK4-binding proteins to generate mimetics, e.g. peptide or non-peptide agents, which are able to mimic binding of the authentic protein to another cellular partner, e.g. a cyclin-dependent kinase, e.g. CDK4, or other cellular protein, e.g., an *erk* kinase, p53 or Src, etc. Such mutagenic techniques as described above, as well as the thioredoxin system, are also particularly useful for mapping the determinants of a CDK4-binding protein which participate in protein-protein interactions involved in, for example, binding of the subject protein to CDK4, CDK6 etc. To illustrate, the critical residues of a CDK4-binding protein which are involved in molecular recognition of CDK4 can be determined and used to generate peptidomimetics which bind to CDK4, and by inhibiting binding of the CDK4-binding protein, act to prevent activation of the kinase. By employing, for example, scanning mutagenesis to map the amino acid residues of the CDK4-binding protein which are involved in binding CDK4, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those

residues in binding to the kinase. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J. Med. Chem.* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β -turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans 1*:1231), and β -aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Another aspect of the invention pertains to an antibody specifically reactive with one of the subject CDK4-binding proteins. For example, by using immunogens derived from the present activity CDK4-binding proteins, based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or a rabbit can be immunized with an immunogenic form of the polypeptide (e.g., CDK4-binding protein or an antigenic fragment which is capable of eliciting an antibody response). Techniques for conferring immunogenicity on a protein or polypeptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject CDK4-binding proteins can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the CDK4-binding proteins of the present invention, e.g. antigenic determinants of a protein represented by one of SEQ ID NOS: 25-48 or a closely related human or non-human mammalian homolog (e.g. 90 percent homologous, more preferably at least 95 percent homologous). In yet a further preferred embodiment of the present invention, the anti-CDK4-BP antibodies do not substantially cross react (i.e. react specifically) with a protein which is: e.g. less than 90 percent homologous to one of SEQ ID NOS: 25-48; e.g. less than 95 percent homologous with one of SEQ ID NOS: 25-48; e.g. less than 98-99 percent homologous with one of SEQ ID NOS: 25-48. By "not substantially cross react", it is meant that the antibody has a binding affinity for a nonhomologous protein (e.g. CDK4) which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity of that antibody for a protein of SEQ ID NOS: 25-48.

Following immunization, anti-CDK4-BP antisera can be obtained and, if desired, polyclonal anti-CDK4-BP antibodies isolated from the serum. To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, and include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) *Nature*, 256: 495-497), the human B cell hybridoma technique (Kozbar *et al.*, (1983) *Immunology Today*, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, (1985) *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened immunochemically for production of antibodies specifically reactive with a CDK4-binding protein of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with one of the subject CDK4-binding protein. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having an anti-CDK4-BP portion.

Both monoclonal and polyclonal antibodies (Ab) directed against the subject CDK4-BP or CDK4-BP variants, and antibody fragments such as Fab' and $F(ab')_2$, can be used to block the action of a subject CDK4-BP and allow the study of the role of a particular CDK4 binding protein of the present invention in the normal cellular function of the subject CDK4-binding protein, e.g. by microinjection of anti-CDK4BP antibodies of the present invention.

Antibodies which specifically bind CDK4-BP epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of each of the subject CDK4-binding protein. Anti-CDK4-BP antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate CDK4-BP levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor CDK4-BP levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with a disorder. The level of CDK4-BP can be measured in cells found in bodily fluid, such as in samples of cerebral spinal fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-CDK4-BP antibodies can include, for example, immunoassays designed to aid in early diagnosis of a neoplastic or hyperplastic disorder, e.g. the presence of cancerous cells in the sample, e.g. to detect cells in which a lesion of the CDK4-BP gene has occurred.

Another application of anti-CDK4-BP antibodies is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λ gt18-23, λ ZAP, and λ ORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λ gt11 will produce
5 fusion proteins whose amino termini consist of β -galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject CDK4-BP can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-CDK4-BP antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of CDK4-BP homologs can
10 be detected and cloned from other sources, and alternate isoforms (including splicing variants) can be detected and cloned from human sources.

Antibodies which are specifically immunoreactive with a CDK4-binding protein of the present invention can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of the protein. Such antibodies can
15 be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate levels of one or more CDK4-binding proteins in tissue or cells isolated from a bodily fluid as part of a clinical testing procedure. For instance, such measurements can be useful in predictive valuations of the onset or progression of tumors. Likewise, the ability to monitor certain CDK4-binding protein levels in an individual can allow determination of the efficacy
20 of a given treatment regimen for an individual afflicted with such a disorder. Diagnostic assays using the subject antibodies, can include, for example, immunoassays designed to aid in early diagnosis of a neoplastic or hyperplastic disorder, e.g. the presence of cancerous cells in the sample, e.g. to detect cells in which alterations in expression levels of a CDK4-BP gene has occurred relative to normal cells.

25 In addition, nucleotide probes can be generated from the cloned sequence of the CDK4-BP genes, which probes will allow for histological screening of intact tissue and tissue samples for the presence of a CDK4-BP-encoding mRNA. Similar to the diagnostic uses of the subject antibodies, the use of probes directed to CDK4-BP messages, or to genomic CDK4-BP gene sequences, can be used for both predictive and therapeutic
30 evaluation of allelic mutations or abnormal transcription which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth).

Accordingly, the present method provides a method for determining if a subject is at risk for a disorder characterized by unwanted cell proliferation. In preferred embodiments, the method can be generally characterized as comprising detection, in a tissue of the subject,
35 the presence or absence of a genetic lesion manifest as at least one of (i) a mutation of a gene encoding a CDK4-binding protein, or (ii) the mis-expression of that gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of

one or more nucleotides from a gene, (ii) an addition of one or more nucleotides to a gene, (iii) a substitution of one or more nucleotides of a gene, (iv) a gross chromosomal rearrangement of a gene, (v) a gross alteration in the level of a messenger RNA transcript of a gene, (vi) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a gene, and (vii) a non-wild type level of a CDK4-binding protein. In one aspect of the invention, there is provided a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which is capable of hybridizing to a sense or antisense sequence of one of SEQ. ID Nos: 1-24, or naturally occurring mutants thereof, or 5' or 3' flanking sequences or intronic sequences naturally associated with the subject CDK4-BP gene or naturally occurring mutants thereof. The probe is exposed to nucleic acid of a tissue sample; and the hybridization of the probe to the sample nucleic acid is detected. In certain embodiments, detection of the lesion comprises utilizing the probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1944) *PNAS* 91:360-364), the later of which can be particularly useful for detecting point mutations in the gene. Alternatively, the level of a CDK4-binding protein can detected in an immunoassay.

As set out above, the present invention also provides assays for identifying drugs which are either agonists or antagonists of the normal cellular function of a CDK4-binding protein, or of the role of that protein in the pathogenesis of normal or abnormal cellular proliferation and/or differentiation and disorders related thereto, as mediated by, for example binding of the CDK4-binding protein to a target protein, e.g., CDK4, CDK6, or another cellular protein. In one embodiment, the assay evaluates the ability of a compound to modulate binding of a CDK4-binding protein to a CDK or other of cell-cycle regulatory protein. While the following description is directed generally to embodiments exploiting the interaction between a CDK4-binding protein, *cdc37*, and a CDK, it will be understood that these examples are merely illustrative, and that similar embodiments can be generated using, for example, a *erk* polypeptide, such as *erk1* or *erk2*, as target proteins for *cdc37*. Moreover, the other CDK4-binding proteins of the present invention can be exploited in similar assays.

A variety of assay formats will suffice and, in light of the present disclosure, those not expressly described herein will nevertheless be comprehended by one of ordinary skill in the art. Agents to be tested for their ability to act as *cdc37* inhibitors can be produced, for example, by bacteria, yeast or other organisms (e.g. natural products), produced chemically (e.g. small molecules, including peptidomimetics), or produced recombinantly. In a preferred embodiment, the test agent is a small organic molecule, e.g., other than a peptide, oligonucleotide, or analog thereof, having a molecular weight of less than about 2,000 daltons.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity between *cdc37* and other proteins, or in changes in a property of the molecular target for *cdc37* binding. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *cdc37* polypeptide which is ordinarily capable of binding CDK4. To the mixture of the compound and *cdc37* polypeptide is then added a composition containing a CDK4 polypeptide. Detection and quantification of CDK4/*cdc37* complexes provides a means for determining the compound's efficacy at inhibiting (or potentiating) complex formation between the CDK4 and *cdc37* polypeptides. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified CDK4 is added to a composition containing the *cdc37* protein, and the formation of CDK4/*cdc37* complex is quantitated in the absence of the test compound. It will be understood that, in general, the order in which the reactants may be admixed can be varied, and can be admixed simultaneously. Moreover, CDK4 can be substituted with other proteins to which *cdc37* binds, as a complex by immunoprecipitation of *cdc37* by anti-*cdc37* antibodies, such as a protein having a molecular weight of approximately 40kd, 42kd, 95kd, 107kd and 117kd.

Complex formation between the *cdc37* polypeptide and target polypeptide may be detected by a variety of techniques. For instance, modulation of the formation of complexes can be quantitated using, for example, detectably labelled proteins such as radiolabelled (e.g. ^{32}P , ^{35}S , ^{14}C or ^3H), fluorescently labelled (e.g. FITC), or enzymatically labelled *cdc37* or CDK4 polypeptides, by immunoassay, or by chromatographic detection. The use of enzymatically labeled CDK4 will, of course, generally be used only when enzymatically inactive portions of CDK4 are used, as each protein can possess a measurable intrinsic activity which can be detected.

Typically, it will be desirable to immobilize either the *cdc37* or the CDK4 polypeptide to facilitate separation of *cdc37*/CDK4 complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of

CDK4 to *cdc37*, in the presence and absence of a candidate agent, can be accomplished in any vessel suitable for containing the reactants. Examples include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/*cdc37* (GST/*cdc37*) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the CDK4 polypeptide, e.g. an ³⁵S-labeled CDK4 polypeptide, and the test compound, and the mixture incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired, e.g., at 4°C in a buffer containing 0.6M NaCl or a detergent such as 0.1% Triton X-100. Following incubation, the beads are washed to remove any unbound CDK4 polypeptide, and the matrix immobilized radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the *cdc37*/CDK4 complexes are subsequently dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of CDK4 polypeptide found in the bead fraction quantitated from the gel using standard electrophoretic techniques such as described in the appended examples.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, either of the *cdc37* or CDK4 proteins can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated *cdc37* molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the *cdc37* but which do not interfere with CDK4 binding can be derivatized to the wells of the plate, and the *cdc37* trapped in the wells by antibody conjugation. As above, preparations of a CDK4 polypeptide and a test compound are incubated in the *cdc37*-presenting wells of the plate, and the amount of *cdc37*/CDK4 complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the CDK4 polypeptide, or which are reactive with the *cdc37* protein and compete for binding with the CDK4 polypeptide; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the CDK4 polypeptide, either intrinsic or extrinsic activity. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with a CDK4 polypeptide. To illustrate, the CDK4 polypeptide can be chemically cross-linked or genetically fused with horseradish peroxidase, and the amount of CDK4 polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. 3,3'-diamino-benzadine tetrahydrochloride or 4-chloro-1-naphthol. Likewise, a fusion protein comprising the CDK4

polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) *J Biol Chem* 249:7130). Direct detection of the kinase activity of CDK4 can be provided using substrates known in the art, e.g., histone H1.

5 For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as either anti-CDK4 or anti-*cdc37* antibodies, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the CDK4 polypeptide or *cdc37* sequence, a second polypeptide for which antibodies are readily
10 available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) *J Biol Chem* 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharmacia, NJ).

15 Moreover, the subject *cdc37* polypeptides can be used to generate an interaction trap assay, as described in the examples below (see also, U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J Biol Chem* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; and Iwabuchi et al. (1993) *Oncogene* 8:1693-1696), for subsequently detecting agents which disrupt binding of *cdc37* to a CDK or other cell-cycle
20 regulatory protein, e.g. Src or p53.

The interaction trap assay relies on reconstituting *in vivo* a functional transcriptional activator protein from two separate fusion proteins, one of which comprises the DNA-binding domain of a transcriptional activator fused to a CDK, such as CDK4. The second fusion protein comprises a transcriptional activation domain (e.g. able to initiate RNA polymerase
25 transcription) fused to a *cdc37* polypeptide. When the CDK4 and *cdc37* domains of each fusion protein interact, the two domains of the transcriptional activator protein are brought into sufficient proximity as to cause transcription of a reporter gene. By detecting the level of transcription of the reporter, the ability of a test agent to inhibit (or potentiate) binding of *cdc37* to CDK4 can be evaluated.

30 In an illustrative embodiment, *Saccharomyces cerevisiae* YPB2 cells are transformed simultaneously with a plasmid encoding a GAL4db-CDK4 fusion and with a plasmid encoding the GAL4ad domain fused to a *cdc37*. Moreover, the strain is transformed such that the GAL4-responsive promoter drives expression of a phenotypic marker. For example, the ability to grow in the absence of histidine can depends on the expression of the HIS3
35 gene. When the HIS3 gene is placed under the control of a GAL4-responsive promoter, relief of this auxotrophic phenotype indicates that a functional GAL4 activator has been

reconstituted through the interaction of CDK4 and the *cdc37*. Thus, a test agent able to inhibit *cdc37* interaction with CDK4 will result in yeast cells unable to grow in the absence of histidine. Alternatively, the phenotypic marker (e.g. instead of the *HIS3* gene) can be one which provides a negative selection (e.g., are cytotoxic) when expressed such that agents which disrupt CDK4/*cdc37* interactions confer positive growth selection to the cells.

In yet another embodiment, a mammalian *cdc37* gene can be used to rescue a yeast cell having a defective *Cdc37* gene, such as the temperature sensitive mutant (*Cdc37*^{TS}; see Reed (1980) *Genetics* 95:561-577; and Reed et al. (1985) *CSH Symp Quant Biol* 50:627-634). For example, a humanized yeast can be generated by amplifying the coding sequence of the human *cdc37* clone, and subcloning this sequence into a vector which contains the yeast *GAL* promoter and *ACT1* termination sequences flanking the *cdc37* coding sequences. This plasmid can then be used to transform a *Cdc37*^{TS} mutant (Gietz et al. (1992) *Nuc Acid Res* 20:1425). To assay growth rates, cultures of the transformed cells can be grown at 37°C (an impermissive temperature for the TS mutant) in media supplemented with galactose. Turbidity measurements, for example, can be used to easily determine the growth rate. At the non-permissive temperature, growth of the yeast cells becomes dependent upon expression of the human *cdc37* protein. Accordingly, the humanized yeast cells can be utilized to identify compounds which inhibit the action of the human *cdc37* protein. It is also deemed to be within the scope of this invention that the humanized yeast cells of the present assay can be generated so as to comprise other human cell-cycle proteins. For example, human CDKs and human cyclins can also be expressed in the yeast cell. To illustrate, a triple *cln* deletion mutant of *S. Cerevisiae* which is also conditionally deficient in *cdc28* (the budding yeast equivalent of *cdc2*) can be rescued by the co-expression of a human cyclin D1 and human CDK4, demonstrating that yeast cell-cycle machinery can be at least in part replaced with corresponding human regulatory proteins. Roberts et al. (1993) *PCT Publication Number WO 93/06123*. In this manner, the reagent cells of the present assay can be generated to more closely approximate the natural interactions which the mammalian *cdc37* protein might experience.

Furthermore, it will be possible to perform such assays as differential screening assays, which permit comparison of the effects of a drug on a number of different complexes formed between the CDK4-binding protein and other cell-cycle regulatory proteins, e.g. other CDKs. For instance, each of the above assays can be run with a subject CDK4-BP and each of CDK4, CDK5 and CDK6. In side-by-side comparison, therefore, agents can be chosen which selectively effect the formation of, for example, the CDK-BP/CDK4 complex without substantially interfering with the other CDK complexes.

Moreover, certain formats of the subject assays can be used to identify drugs which inhibit proliferation of yeast cells or other lower eukaryotes, but which have a substantially

reduced effect on mammalian cells, thereby improving therapeutic index of the drug as an anti-mycotic agent. To illustrate, the identification of such compounds is made possible by the use of differential screening assays which detect and compare drug-mediated disruption of binding between two or more different types of *cdc37*/CDK complexes. Differential screening assays can be used to exploit the difference in drug-mediated disruption of human CDK/*cdc37* complexes and yeast CDC2/*Cdc37* complexes in order to identify agents which display a statistically significant increase in specificity for disrupting the yeast complexes relative to the human complexes. Thus, lead compounds which act specifically to inhibit proliferation of pathogens, such as fungus involved in mycotic infections, can be developed. By way of illustration, the present assays can be used to screen for agents which may ultimately be useful for inhibiting at least one fungus implicated in such mycosis as candidiasis, aspergillosis, mucormycosis, blastomycosis, geotrichosis, cryptococcosis, chromoblastomycosis, coccidioidomycosis, conidiosporosis, histoplasmosis, maduromycosis, rhinosporidiosis, nocardiosis, para-actinomycosis, penicilliosis, moniliasis, or sporotrichosis. For example, if the mycotic infection to which treatment is desired is candidiasis, the present assay can comprise comparing the relative effectiveness of a test compound on mediating disruption of a human CDK4/*cdc37* complex with its effectiveness towards disrupting the equivalent complexes formed from genes cloned from yeast selected from the group consisting of *Candida albicans*, *Candida stellatoidea*, *Candida tropicalis*, *Candida parapsilosis*, *Candida krusei*, *Candida pseudotropicalis*, *Candida quillermondii*, or *Candida rugosa*. Likewise, the present assay can be used to identify anti-fungal agents which may have therapeutic value in the treatment of aspergillosis by making use of an interaction trap assays derived from CDK and *Cdc37* genes cloned from yeast such as *Aspergillus fumigatus*, *Aspergillus flavus*, *Aspergillus niger*, *Aspergillus nidulans*, or *Aspergillus terreus*. Where the mycotic infection is mucormycosis, the complexes can be derived from yeast such as *Rhizopus arrhizus*, *Rhizopus oryzae*, *Absidia corymbifera*, *Absidia ramosa*, or *Mucor pusillus*. Sources of other *Cdc37*-containing complexes for comparison with a human CDK/*cdc37* complex includes the pathogen *Pneumocystis carinii*.

Moreover, inhibitors of the enzymatic activity of any of the subject CDK-binding proteins which are enzymes, e.g. a kinase, e.g. an isopeptidase, e.g. a protease, can be identified using assays derived from measuring the ability of an agent to inhibit catalytic conversion of a substrate by the subject proteins.

In another aspect, the invention features transgenic non-human animals which express a recombinant CDK4-BP gene of the present invention, or which have had one or more of the subject CDK4-BP gene(s), e.g. heterozygous or homozygous, disrupted in at least one of the tissue or cell-types of the animal.

In another aspect, the invention features an animal model for developmental diseases, which has a CDK4-BP allele which is mis-expressed. For example, a mouse can be bred which has a CDK4-BP allele deleted, or in which all or part of one or more CDK4-BP exons are deleted. Such a mouse model can then be used to study disorders arising from mis-expressed CDK4-BP genes.

Exemplification

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Interaction Trap

A general transcription-based selection for protein-protein interactions was used to isolate cDNA which encode proteins able to bind to CDK4. Development of the "interaction trap assay" or ITS, is described in, for example, Gyuris et al. (1993) *Cell* 75:791-803; Chien et al. (1991) *PNAS* 88:9578-9582; Dalton et al. (1992) *Cell* 68:597-612; Durfee et al. (1993) *Genes Dev* 7:555-569; Vojtech et al. (1993) *Cell* 74:205-214; Fields et al. (1989) *Nature* 340:245-246; and U.S. Patent Serial number 5,283,173). As carried out in the present invention, the interaction trap comprises three different components: a fusion protein that contains the LexA DNA-binding domain and that is known to be transcriptionally inert (the "bait"); reporter genes that have no basal transcription and whose transcriptional regulatory sequences are dependent on binding of LexA; and the proteins encoded by an expression library, which are expressed as chimeras and whose amino termini contain an activation domain and other useful moieties (the "fish"). Briefly, baits were produced constitutively from a 211 HIS3+ plasmid under the control of the ADH1 promoter and contained the LexA carboxy-terminal oligomerization region. Baits were made in pLexA(1-202)+pl (described in Ruden et al. *Nature* (1991) 350:250-252; and Gyuris et al. *Cell* (1993) 75:791-803) after PCR amplification of the bait coding sequences from the second amino acid to the Stop codon, except for p53 where the bait moiety starts at amino acid 74. Using the PCR primers described in Table I, CDK2 and CDK3 were cloned as EcoR1-BamH1 fragments; CDK4, cyclin D1, cyclin D2, Cyclin E as EcoR1-Sal1 fragments; CDK5, CDK6, Cdi1 as EcoR1-Xho1 fragments; and retinoblastoma (pRb), mutRb(Δ 702-737), p53 and cyclin C as BamH1-Sal1 fragments. When EcoR1 is used, there are two amino acid inserted (EF) between the last amino acid of LexA and the bait moieties. BamH1 fusion results in five amino acid insertion (EFPGI) between LexA and the fused protein.

PCR primers:

CDK2:

5' -GGCGGCCGCGAATTCGAGAAGCTTCCAAAAGGTGGAAAAG-3'
5' -GCGGCCGCGGATCCAGGCTATCAGAGTCGAAGATGGGGTAC-3'

CDK3:

5' -GCGGCCGCGAATTCGAAGCTGGAGGAGCAACCGGGAGC-3'
5' -GCGGCCGCGGATCCTCAATGGCGGAATCGCTGCAGCAC-3'

10

CDK5:

5' -GCGGCGGCGTCGACCAGAAATACGAGAACTGGAAAAG-3'
5' -GCGGCGGCGTCGACCGGGGCCTAGGGCGGACAGAAGTC-3'

15 CDK6:

5' -GCGGCCGCGAATTCGAGAAGGACGGCCTGTGCCGCGCT-3'
5' -GCGGCGGCCTCGAGGAGGCCTCAGGCTGTATTAGCTC-3'

Cyclin C:

20 5' -GGCCGGCCGGGATCCTTGTGCTCCGCGGCTGCTCCGGCTG-3'
5' -GCGGCCGCGTCGACGTTTAAAGATTGGCTGTAGCTAGAG-3'

Cyclin D1:

25 5' -GGCCGGCCGGAATTCGAACACCAGCTCCTGTGCTGCGAAG-3'
5' -GCGGCCGCGTCGACGCGCCCTCAGATGTCCACGTCCCCG-3'

Cyclin D2:

5' -GCGGCGGCGAATTCGAGCTGCTGTGCCACGAGGTGGAC-3'
5' -GCGGCGGCGAATTCGAGCTGCTGTGCCACGAGGTGGAC-3'

30

Cyclin E:

5' -GGCCGGCCGGAATTCAGGAGGACGGCGGCGGAGTTC-3'
5' -GCGGCCGCGTCGACGGGTGGTCACGCCATTTCCGGCCCG-3'

35 Cdi1:

5' -GCGGCCGCGAATTCAGCCGCCAGTTCAATACAAACAAG-3'
5' -GCGGCCGCCTCGAGATTCCTTTATCTTGATACAGATCTTG-3'

Rb:

40 5' -GCGGCCGCGGATCCAGCCGCCAAAACCCCCGAAAAACG-3'
5' -GCGGCCGCGAATTCCTCGAGCTCATTTCTCTTCTTGTGAGG-3'

p53:

45 5' -GCGGCCGCGGATCCAAGCCCCTGCACCAGCAGCTCCTACA-3'
5' -GCGGCCGCGTCGACTCAGTCTGAGTCAGGCCCTTCTGT-3'

Reporters

The LexAop-LEU2 construction replaced the yeast chromosomal LEU2 gene. The other reporter, pRB1840, one of a series of LexAop-GAL1-lacZ genes (Brent et al. (1985) *Cell* 43:729-736; Kamens et al. (1990) *Mol Cell Biol* 10:2840-2847), was carried on a 2 μ plasmid. Basal reporter transcription was extremely low, presumably owing both to the removal of the entire upstream activating sequence from both reporters and to the fact that LexA operators introduced into yeast promoters decrease their transcription (Brent and Ptashne (1984) *Nature* 312:612-615). Reporters were chosen to differ in sensitivity. The LEU2 reporter contained three copies of the high affinity LexA-binding site found upstream of *E. coli* colE1, which presumably bind a total of six dimers of the bait. In contrast, the lacZ gene contained a single lower affinity operator that binds a single dimer of the bait. The operators in the LEU2 reporter were closer to the transcription start point than they were in the lacZ reporter. These differences in the number, affinity, and operator position all contribute to that fact that the LEU2 reporter is more sensitive than the lacZ gene.

15 Expression Vectors and Library

Library proteins were expressed from pJG4-5, a member of a series of expression plasmids designed to be used in the interaction trap and to facilitate analysis of isolated proteins. These plasmids carry the 2 μ replicator and the TRP1 marker. pJG4-5, shown in Figure 1, directs the synthesis of fusion proteins. Proteins expressed from this vector possess the following features: galactose-inducible expression so that their synthesis is conditional, an epitope tag to facilitate detection, a nuclear localization signal to maximize intranuclear concentration to increase selection sensitivity, and an activation domain derived from *E. coli* (Ma and Ptashne (1987) *Cell* 57:113-119), chosen because its activity is not subject to known regulation by yeast proteins and because it is weak enough to avoid toxicity (Gill and Ptashne (1988) *Nature* 334:721-724; Berger et al. (1992) *Cell* 70:251-265) that might restrict the number or type of interacting proteins recovered. We introduced EcoRI-XhoI cDNA-containing fragments, which were generated from a quiescent normal fibroblast line (WI38), into the pJG4-5 plasmid.

30 CDK4 Interaction Trap

We began with yeast cells which contained LexAop-LEU2 and LexAop-lacZ reporters and the LexA-CDK4 bait. We introduced the WI38 cDNA library (in pJG4-5) into this strain. We recovered a number of transformants on glucose Ura⁻ His⁻ Trp⁻ plates, scraped them, suspended them in approximately 20 ml of 65% glycerol, 10 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, and stored the cells in 1 ml aliquots at -80°C. We determined plating efficiency on galactose Ura⁻ His⁻ Trp⁻ after growing 50 μ l of cell suspension for 5 hr in 5 ml of YP medium, 2% galactose. For the selection, about 2×10^7 galactose-viable cells were plated on four standard circular 10 cm galactose Ura⁻ His⁻ Trp⁻ Leu⁻ plates after galactose

induction. After 4 days at 30°C, LEU⁺ colonies appeared and were collected on glucose Ura⁻ His⁻ Trp⁻ master plates and retested on glucose Ura⁻ His⁻ Trp⁻ Leu⁻, galactose Ura⁻ His⁻ Trp⁻ Leu⁻, glucose X-Gal Ura⁻ His⁻ Trp⁻, and galactose X-Gal Ura⁻ His⁻ Trp⁻ plates. Of these, plasmid DNAs were rescued from colonies which showed galactose-dependent growth on
5 Leu⁻ media and galactose-dependent blue color on X-Gal medium (Hoffman and Winston, (1987) *Gene* 57:267-272), introduced into *E. coli* KC8, and transformants collected on Trp⁻ ampicillin plates.

We classified library plasmids by restriction pattern on 1.8% agarose, 0.5 x Tris-borate-EDTA gels after digestion with EcoRI and XhoI and either AluI or HaeIII. We
10 reintroduced those plasmids from each map class that contained the longest cDNAs into EGY48 derivatives that contained a panel of different baits, e.g. other CDKs, cyclins, p53, Rb, etc. As is evident from inspection of the data for this experiment (see Figure 2), each of the subject CDK4-binding proteins displayed different binding affinities for other cell-cycle regulatory proteins. This finding is significant for a number of reasons. For example, in
15 choosing a particular CDK4 interaction as a therapeutic target for drug design, therapeutic index concerns might cause selection of a CDK4-BP target which interacts primarily with CDK4 and much less with any other CDK. Alternatively, if desired, the ability of a particular CDK4-BP to bind multiple CDKs can be exploited in testing compounds in differential screening assays as described above. Thus, drugs which can alter the binding of, for
20 example, a particular CDK4-BP to CDK4 but which have less effect on the same complex formed with CDK5, will presumably have a better therapeutic index with regard to neuronal side effects than a drug which interferes equally with both.

Furthermore, a deposit of each of these clones as a library of pJG4-5 plasmids (designated "pJG4-5-CDKBP") containing 24 different proteins isolated in the CDK4
25 interaction trap has been made with the American Type Culture Collection (Rockville, MD) on May 26, 1994, under the terms of the Budapest Treaty. ATCC Accession number 75788 has been assigned to the deposit. The cDNAs were inserted into this vector as EcoRI-XhoI fragments. The EcoRI adaptor sequence is 5'-GAATTCTGCGGCCGC-3' and the open reading frame encoding the interacting protein starts with the first G. With this deposit in
30 hand, one of ordinary skill in the art can generate the subject recombinant CDK4-BP genes and express recombinant forms of the subject CDK4-binding proteins. For instance, each of the CDK4-binding proteins of the present invention can be amplified from ATCC deposit no. 75788 by PCR using the following primers:

5'-TAC CAG CCT CTT GCT GAG TGG AGA-3' (SEQ ID No. 71)
35 5'-TAG ACA AGC CGA CAA CCT TGA TTG-3' (SEQ ID No. 72)

Moreover, it will be immediately evident to those skilled in the art that, in light of the guide to the 5' and 3' ends to each of the clones provided in Table 1, each individual clone of the ATCC deposit can be isolated using primers based on the nucleotide sequences provided

by SEQ ID Nos. 1-24 and 49-70, or a combination of such primers and the primers of SEQ ID Nos. 71 and 72.

Isolated clones can be subcloned into expression vectors in order to produce a recombinant protein, or can be used to generate anti-sense constructs, or can be used to generate oligonucleotide probes. In an illustrative embodiment, oligonucleotide probes have been generated using the coding sequences for each of the clones of the subject ATCC deposit, and used in Southern hybridization and *in situ* hybridization assays to detect the pattern and abundance of expression of each of the CDK4-binding proteins.

Moreover, because each member of the ATCC deposit is a plasmid encoding a fusion protein identified from an interaction trap assay, the clone can be utilized directly from the deposit in a similar ITS employed as, for example, a drug screening assay, or alternatively, a mutagenesis assay for mapping CDK4 binding epitopes.

Table 1
Guide to pJG4-5-CDKBP

Clone	Nucleotide	Peptide
11	SEQ ID No. 1	SEQ ID No. 25
13	SEQ ID No. 2	SEQ ID No. 26
22	SEQ ID No. 3	SEQ ID No. 27
36	SEQ ID No. 4 (5') SEQ ID No. 49 (3')	SEQ ID No. 28 (N-terminal)
61	SEQ ID No. 5 (5') SEQ ID No. 50 (3')	SEQ ID No. 29 (N-terminal)
68	SEQ ID No. 6 (5') SEQ ID No. 51 (3')	SEQ ID No. 30 (N-terminal)
71	SEQ ID No. 7 (full length) SEQ ID No. 69 (5') SEQ ID No. 70 (3')	SEQ ID No. 31
75	SEQ ID No. 8 (5') SEQ ID No. 52 (3')	SEQ ID No. 32 (N-terminal)
116	SEQ ID No. 9 (full length) SEQ ID No. 67 (5') SEQ ID No. 68 (3')	SEQ ID No. 33
118	SEQ ID No. 10 (5') SEQ ID No. 55 (3') SEQ ID No. 55 (Internal)	SEQ ID No. 34 (N-terminal)
121	SEQ ID No. 11 (5') SEQ ID No. 56 (3')	SEQ ID No. 35 (N-terminal)
125	SEQ ID No. 12 (5') SEQ ID No. 57 (3')	SEQ ID No. 36 (N-terminal)
127	SEQ ID No. 13	SEQ ID No. 37
166	SEQ ID No. 15	SEQ ID No. 39

190	SEQ ID No. 16 (5') SEQ ID No. 58 (3')	SEQ ID No. 40 (N-terminal)
193	SEQ ID No. 17	SEQ ID No. 41
216	SEQ ID No. 18 (5') SEQ ID No. 59 (3')	SEQ ID No. 42
225	SEQ ID No. 19	SEQ ID No. 43
227	SEQ ID No. 20 (5') SEQ ID No. 61 (3')	SEQ ID No. 44 (N-terminal)
267	SEQ ID No. 21	SEQ ID No. 45
269	SEQ ID No. 22 (5') SEQ ID No. 63 (3')	SEQ ID No. 46 (N-terminal)
295	SEQ ID No. 23 (5') SEQ ID No. 64 (3')	SEQ ID No. 47 (N-terminal)

All of the above-cited references and publications are hereby incorporated by reference.

5

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

10

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Mitotix, Inc.
(B) STREET: One Kendall Square, Building 600
(C) CITY: Cambridge
(D) STATE: MA
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 02139
(G) TELEPHONE: (617) 225-0001
(H) TELEFAX: (617) 225-0005

(ii) TITLE OF INVENTION: CDK4-Binding Proteins

(iii) NUMBER OF SEQUENCES: 72

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: ASCII (text)

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/253,155
(B) FILING DATE: 2-JUN-1994

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1638 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

45	GAATTCTGCG GCCGCATGGA TACAGATACA GATACATTCA CCTGTCAGAA AGATGGTCCG	60
	TGGTTCCCTG AGAGAATCTC CTGCAGTCCT AAAAAATGTC CTCTCCCGGA AAACATAACA	120
	CATATACTTG TACATGGGGA CGATTTCAGT GTGAATAGGC AAGTTTCTGT GTCATGTGCA	180
50	GAAGGGTATA CCTTTGAGGG AGTTAACATA TCAATATGTC AGCTTGATGG AACCTGGGAG	240
	CCACCATCTT CCGATGAATC TTGCAGTCCA GTTTCTTGTG GGAAACCAGA AAGTCCAGAA	300
55	CATCGATTTG TGGTTGGCAG TAAATACACC TTTGCAAAAGC ACAATTATTT ATCAGTGTGA	360
	GCCTGGCTAT GAACTGGAGG GGAACAGGGC AACGTGTCTG CCAGGAGAAC AGACAGTGGA	420

	GTGGAGGGGT GGCAATATGC AAAGAGACCA GGTGTGAAAC TCCACTTGAA TTTCTCAATG	480
	GGAAAGCTGA CATTGAAAAC AGGACGACTG GACCCAACGT GGTATATTCC TGCAACAGAG	540
5	GCTACAGTCT TGAAGGGCCA TCTGAGGCAC ACTGCACAGA AAATGGAACC TGGAGCCACC	600
	CAGTCCCTCT CTGCAAACCA AATCCATGCC CTGTTCTTTT TGGTGATTCC CGAGAATGCT	660
	CTGCTGTCTT GAAAAGGAGT TTTATGTTGA TCAGAATGTG TCCATCAAAT GTAGGGAAGG	720
10	TTTTCTGCTG CAGGGCCACG GCATCATTAC CTGCAACCCC GACGAGACGT GGACACAGAC	780
	AAGCGCCAAA TGTGAAAAAA TCTCATGTGG TCCACCAGCT CACGTAGCAA AATGCAATTG	840
15	CTCGAGGCGT ACATTATCAA TATGGAGACA TGATCACCTA CTCATGTTAC AGTGGATACA	900
	TGTTGGAGGG TTTCCTGAGG AGTGTGTTGT TAGAAAATGG AACATGGACA TCACCTCCTA	960
	TTTGACAGAGC TGTCTGTGCGA TTTCCATGTC AAGAATGGGG GCATCTGCCA ACGCCCAAAT	1020
20	GCTTGTTCTT GTCCAGAGGG CTGGATGGGG CGCCTCTTGT GAAGAACCAA TCTGCATTCT	1080
	TCCCTGTCTG AACGGAGGTC GCTGTGTGGC CCCTTACCAG TGTGACTGCC CGCCTGGCTG	1140
25	GACGGGGTCT CGCTGTCAAA CAAGCTGTTT GCCAGTCTCC CTGCTTAAAT GGTGGAAAAT	1200
	GTGTAAGACC AAACCGATGT CACTGCTTTT CTTCTTGGAC GGGACATAAC TGTTCAGGA	1260
	AAAGGAGGAC TGGGTTTTAA CCACTGCACG ACCATCTGGC TCTCCCCAAA GCAGGATCAT	1320
30	CTCTCCTCGG TAGTGCCTGG GCATCCTGGA ACTTATGCGA AGAAAGTCCA ACATGGTGCT	1380
	GGGTCTTGTT TAGTAACTT GTTACTTGGG GTTACTTTTT TTATTTTGTG ATAAATTTTG	1440
35	TTATTCCTTG TGACAACTT TCTTACATGT TTCCATTTTT AAATATGCCT GTATTTTCTA	1500
	AATAAAATT ATATTAAATA GATGCTGCTC TACCCTCACC AAATGTACAT ATTCTGCTGT	1560
	CTATTGGGAA AGTTCCTGGT ACACATTTTT ATTCAATTAC TTAAATGAT TTTTTCATT	1620
40	AAAGTATATT TTGCTACT	1638

(2) INFORMATION FOR SEQ ID NO:2:

- 45 (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 794 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: cDNA

55 (2) INFORMATION FOR SEQ ID NO:2:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 791 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

10 GAATTCTGCG GCCGCGAACT GCTGGCTGCC CACGGTACTC TGGAGCTGCA AGCCGAGATC 60
CTGCCCCGCC GGCCTCCAC GCCGGAGGCC CAGAGCGAAG AGGAGAGATC CGATGAGGAG 120
15 CCGGAGGCCA AAGAAGAGGA AGAGGAAAAA CCACACATGC CCACGGAATT TGATTTTGAT 180
GATGAGCCAG TGACACCAA GGACTCCCTG ATTGACCGGA GACGCACCCC AGGAAGCTCA 240
GCCCCGAGCC AGAAACGGGA GGCCCCCTG GACAAGGTGC TGTCGGACAT GAAGAGACAC 300
20 AAGAAGCTGG AGGAGCAGAT CCTTCGTACC GGGAGGGACC TCCTCAGCCT GGACTCGGAG 360
GACCCAGCC CCGCCAGCCC CCCACTCCGA TCCTCCGGGA GTAGTCTCTT CCCTCGGCAG 420
25 CGGAAATACT GATTCCCACT GCTCCTGCCT CTAGGGTGCA GTGTCCGTAC CTGCTGGAGC 480
CTGGGCCCTC CTCCCCAGC CCAGACATTG AGAACTTG GAAGAAGAGA GAAACCTCAA 540
GCTCCCAAAC AGCACGTTGC GGGAAAGAGG AAGAGAGAGT GTGAGTGTGT GTGTGTGTTT 600
30 TTTCTATTGA ACACCTGTAG AGTGTGTGTG TGTGTTTTCT ATTGAACACC TATAGAGAGA 660
GTGTGTGTGT TTTCTATTGA ACATCTATAT AGAGAGAGTG TGTGAGTGTG TGTGTTCTAT 720
35 TGGACACCTA TTCAGAGACC TGGACTGGAT TTTCTGAGTC TGAAATAAAA GATGCAGAGC 780
TATCATCTCT T

(2) INFORMATION FOR SEQ ID NO:3:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 795 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
45 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATTCTGCG GCCGCGTGGG GACTGAGGAG GATGGCGGAG GCGTCGGCCA CAGGACGGTG 60
55 TACTTGTGTTG ATCGGCGCGA AAAGGAGTCC GAGCTCGGGG ACCGGCCTCT GCAGGTCGGG 120
GAGCGCTCGG ACTACGCGGG ATTCGCGCG TGTGTGTGTC AGACACTTGG CATTTCACCT 180

GAAGAAAAAT TTGTTATTAC AACAAACAAGT AGGAAAGAAA TTACCTGTGA TAATTTTGAT 240
 GAAACTGTTA AAGATGGAGT CACCTTATAC CTGCTACAGT CGGTCAATCA GTTACTACTG 300
 5 ACAGCTACGA AAGAACGAAT TGAATTCTTA CCTCACTATG ACACACTGGT TAAAAGTGGC 360
 ATGTATGAAT ATTATGCCAG TGAAGGACAA AATCCTTTGC CATTTGCTCT TGC GGAATTA 420
 ATTGACAATT CATTGTCTGC TACTTCTCGT AACATTGGGG TTAGAAGAAT ACAGATCAAA 480
 10 TTGCTTTTTG ATGAAACACA AGGAAAACCT GCTGTTGCAG TGATAGATAA TGGAAGAGGA 540
 ATGACCTCTA AACAGCTTAA CAACTGGGCC GTGTATAGGT TGTCAAAATT CACAAGGCAA 600
 15 GGTGACTTTG AAAGTGATCA TTCAGGATGT TCGTCCAGTA CCAGTGCCAC GCAGTTTAAA 660
 TAGTGATATT TCCTATTTGG GTGTTGGGGG CAAGCAAGCT GTCTTCTTTG GTTGGGACAA 720
 TCAGCCAGAA TGATAAGCCA ACCTGCAGAT TCCCCAGATG TTCACGAGCT TGTGCTTTGC 780
 20 TAAAGGAGAT TTTGG 795

(2) INFORMATION FOR SEQ ID NO:4:

25

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 305 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 30 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCTGCG GCCGCGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 60
 40 AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 120
 AGAGAGAGAG AGAGAGAGAG AGAGAGCATT CGGCCCCGATA TGTCTCGCTC CGTGGCCTTA 180
 GATGTTCTCG CTCTACTCTC TCTCTTTGC CTGGAGGCTA TCCAGGTTGC TCCCATAGAT 240
 45 TCATGACCTC TCACCTTCTC CAAGAGATTT GGGTGCAACC AAATTGCCGG GATCCAATCT 300
 TTTCC 305

50 (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 305 base pairs
 (B) TYPE: nucleic acid
 55 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

5 GAATTCTGCG GCCGCCTGCC CCACAACTTT CTCACGGTGG CGCCTGGACA CAGTAGTCAC 60
CACAGTCCAG GCCTGCAGGG CCAGGGTGTG ACCCTGCCCC GGGAGCCACC CCTCCCTGAG 120
10 AAGAAGCGGG TCTCGGAGGG GGATCGTTCT TTGGTTTCAG TCTCTCCCTC CTCCAGTGGT 180
TTCTCCAGCC CGCACAGCGG GAGCAACATC AGTATCCCCT TCCCATATGT CCTTCCCGAC 240
15 TTTTCCAAGG CTTCAGAAGG GGGCTCAACT CTGCAGATTG TCCAGGTGAT AAACCTGTGA 300
TCGGG 305

(2) INFORMATION FOR SEQ ID NO:6:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 424 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

30 GAATTCTGCG GCCGCCGCCG TCCTCCGGCT GACAGGGGGA GGAGCCCGCC GGGAGGGCCG 60
GGGTCTCGGA CTGGGGAGCC GGGACGGGAG AGCAGCGCAG CCGGGTGCAC CGCGGCCGCG 120
35 CCCCAGGAGG GCTGTTCGGG TCAGCGCCCA CCGCTGCTCC GCGCTGACAG CGCCGGACTG 180
GGGCGGTGCG GGGGGCTTTG CAGGCCGCCA GTGTCGACAT ACTGCTGGAG GAGGTTCCGC 240
40 CCGCGACCCG CTGAGTGGGG CGGCGGCCCG GGGCGACGTA CAGGAGGTTT CGCCGTCTTT 300
CTGCAACCCC CGATTTTGTT GTCATCCCCG ACGGCCCTCC AACCCTCTTT CGATAATCTA 360
CGGTGTCTTC CAAGCTCAAT TCACTGTTTT GGCAAGCAAC CCCCATTCC CCCCTGTAG 420
45 CTTG 424

(2) INFORMATION FOR SEQ ID NO:7:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3407 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
55 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

5	GGCGAGCACT GGCTACGTGC GACTGTGGGG AGCGGCGCGG TGCTGGGTGC TGCGGCGGGC	60
	GATGCTGGCC GCCGCCGGGG GCGGGGTTCC CACTGCAGCA GGAGCGTGGT TGCTCCGAGG	120
	CCAGCGGACC TGCGACGCCT CTCCTCCTTG GGCAGTGTGG GGCCGAGGCC CGGCAATTGG	180
10	GGGCCAATGG CGGGGGTTTT GGGAAGCGAG CAGCCGCGGC GGAGGCGCAT TCTCGGGGGG	240
	CGAGGACGCC TCCGAGGGCG GCGCGGAGGA AGGAGCCGGC GGCGCGGGGG GCAGCGCGGG	300
15	CGCCGGGGAA GGCCCGGTCA TAACGGCGCT CACGCCCATG ACGATCCCCG ATGTGTTTCC	360
	GCACCTGCCG CTCATCGCCA TCACCCGCAA CCCGGTGTTT CCGCGCTTTA TCAAGATTAT	420
	CGAGGTTAAA AATAAGAAGT TGTTGAGCT GCTGAGAAGG AAAGTTCGTC TCGCCAGCC	480
20	TTATGTCGGC GTCTTTCTAA AGAGAGATGA CAGCAATGAG TCGGATGTGG TCGAGAGCCT	540
	GGATGAAATC TACCACACGG GGACGTTTGC CCAGATCCAT GAGATGCAGG ACCTTGGGGA	600
25	CAAGCTGCGC ATGATCGTCA TGGGACACAG AAGAGTCCAT ATCAGCAGAC AGCTGGAGGT	660
	GGAGCCCGAG GAGCCGGAGG CGGAGAACAA GCACAAGCCC CGCAGGAAGT CAAAGCGGGG	720
	CAAGAAGGAG GCGGAGGACG AGCTGAGCGC CAGGCACCCG GCGGAGCTGG CGATGGAGCC	780
30	CACCCCTGAG CTCCCGGCTG AGGTGCTCAT GGTGGAGGTA GAGAACGTTG TCCACGAGGA	840
	CTTCCAGGTC ACGGAGGAGG TGAAGCCCT GACTGCAGAG ATCGTGAAGA CCATCCGGGA	900
35	CATCATTGCC TTGAACCCTC TCTACAGGGA GTCAGTGCTG CAGATGATGC AGGCTGGCCA	960
	GCGGGTGGTG GACAACCCCA TCTACCTGAG CGACATGGGC GCCGCGCTCA CCGGGGCCGA	1020
	GTCCCATGAG CTGCAGGACG TCCTGGAAGA GACCAATATT CCTAAGCGGC TGTACAAGGC	1080
40	CCTCTCCCTG CTGAAGAAGG AATTTGAACT GAGCAAGCTG CAGCAGCGCC TGGGGCGGGA	1140
	GGTGGAGGAG AAGATCAAGC AGACCCACCG TAAGTACCTG CTGCAGGAGC AGCTAAAGAT	1200
45	CATCAAGAAG GAGCTGGGCC TGGAGAAGGA CGACAAGGAT GCCATCGAGG AGAAGTTCCG	1260
	GGAGCGCCTG AAGGAGCTCG TGGTCCCCAA GCACGTCTATG GATGTTGTGG ACGAGGAGCT	1320
	GAGCAAGCTG GGCCTGCTGG ACAACCACTC CTCGGAGTTC AATGTCACCC GCAACTACCT	1380
50	AGACTGGCTC ACGTCCATCC CTTGGGGCAA GTACAGCAAC GAGAACCTGG ACCTGGCGCG	1440
	GGCACAGGCA GTGCTGGAGG AAGACCACTA CGGCATGGAG GACGTCAAGA AACGCATCCT	1500
55	GGAGTTTATT GCCGTTAGCC AGCTCCGCGG CTCCACCCAG GGCAAGATCC TCTGCTTCTA	1560
	TGGCCCCCTT GCGGTGGGTA AGACCAGCAT TGCTCGCTCC ATCGCCCGCG CCCTGAACCG	1620

	AGAGTACTTC CGCTTCAGCG TCGGGGGCAT GACTGACGTG GCTGAGATCA AGGGCCACAG	1680
	GCGGACCTAC GTGGGCGCCA TGCCCGGGAA GATCATCCAG TGTTTGAAGA AGACCAAGAC	1740
5	GGAGAACCCC CTGATCCTCA TCGACGAGGT GGACAAGATC GGCCGAGGCT ACCAGGGGGA	1800
	CCCGTCGTCG GCACTGCTGG AGCTGCTGGA CCCAGAGCAG AATGCCAACT TCCTGGACCA	1860
	CTACCTGGAC GTGCCCCTGG ACTTGTCCAA GGTGCTGTTC ATCTGCACGG CCAACGTAC	1920
10	GGACACCATC CCCGAGCCGC TGCAGACCG TATGGAGATG ATCAACGTGT CAGGCTACGT	1980
	GGCCCAGGAG AAGCTGGCCA TTGCGGAGCG CTACCTGGTG CCCAGGCTC GCGCCCTGTG	2040
15	TGGCTTGGAT GAGAGCAAGG CCAAGCTGTC ATCGGACGTG CTGACGCTGC TCATCAAGCA	2100
	GTA CTGCCG GAGAGCGGTG TCCGCAACCT GCAGAAGCAA GTGGAGAAGG TGTTACGGAA	2160
	ATCGGCCTAC AAGATTGTCA GCGGCGAGGC CGAGTCCGTG GAGGTGACGC CCGAGAACCT	2220
20	GCAGGACTTC GTGGGAAGC CCGTGTTCAC CGTGGAGCGC ATGTATGACG TGACACCGCC	2280
	CGGCGTGGTC ATGGGGCTGG CCTGGACCGC AATGGGAGGC TCCACGCTGT TTGTGGAGAC	2340
25	ATCCCTGAGA CGGCCACAGG ACAAGGATGC CAAGGGTGAC AAGGATGGCA GCCTGGAGGT	2400
	GACAGGCCAG CTGGGGGAGG TGATGAAGGA GAGCGCCCGC ATAGCCTACA CCTTCGCCAG	2460
	AGCCTTCCTC ATGCAGCACG CCCCCGCCAA TGA CTACCTG GTGACCTCAC ACATCCACCT	2520
30	GCATGTGCCC GAGGGCGCCA CCCCCAAGGA CGGCCCAAGC GCAGGCTGCA CCATCGTCAC	2580
	GGCCCTGCTG TCCCTGGCCA TGGGCAGGCC TGTCCGGCAG AATCTGGCCA TGA CTGGCGA	2640
35	AGTCTCCCTC ACGGGCAAGA TCCTGCCTGT TGGTGGCATC AAGGAGAAGA CCATTGCGGC	2700
	CAAGCGCGCA GGGGTGACGT GCATCATCCT GCCAGCCGAG AACAAGAAGG ACTTCTACGA	2760
	CCTGGCAGCC TTCATCACCG AGGGCCTGGA GGTGCACTTC GTGGAACACT ACCGGGAGAT	2820
40	CTTCGACATC GCCTTCCCGG ACGAGCAGGC AGAGGCGCTG GCCGTGGAAC GGTGACGGCC	2880
	ACCCCGGGAC TGCAGGCGGC GGATGTCAGG CCCTGTCTGG GCCAGAACTG AGCGCTGTGG	2940
45	GGAGCGCGCC CGGACCTGGC AGTGGAGCCA CCGAGCGAGC AGCTCGGTCC AGTGACCCAG	3000
	ATCCCAGGGA CCTCAGTCGG CTTAATCAGA GTGTGGCATA GAAGCTATTT AATGATTAAA	3060
	GTCATTTGCA GTGGGAGTTA GCATCACTAA CCTGACAGTT GTTGCCAGGA ATTTGCTTTG	3120
50	TTTACTGCTA GTATATTAGA AATCCTAGAT CTCAGAATCA CAATAGTAAT AAACAACAGG	3180
	GGTCATTTTT TCCTAACTTA CTCTGTGTTT AGGTGTGGAA TTTCTGTCTC CCAAGAGGAA	3240
55	ATGTGACTTC ACTTTGGTGC CAATGGACAG AAAATTCTAC CTGTGCTACA TAGGAGAAGT	3300
	TTGGAATGCA CTTAATAGCT GGTTTTTTACA CCTTGATTTC GAGGTGGAAA GAAATTGATC	3360

ATGAATCTCT AATAAATTTA AATCTCTTAA ACCAAAAAAA AAAAAA

3407

(2) INFORMATION FOR SEQ ID NO:8:

- 5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 450 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAATTCTGCG GCCGCACTGG AGAACCTGC TGTGACTGGG TGGGAGATGA GGGAGCAGGC 60
CACTTCGTGA AGATGGTGCA CAACGGGATA GAGTATGGGG ACATGCAGCT GATCTGTGAG 120
20 GCATACCACC TGATGAAAGA CGTGCTGGGC ATGGCGCAGG ACGAGATGGC CCAGGCCTTT 180
GAGGATTGGA ATAAGACAGA GCTAGACTCA TTCCTGATTG AAATCACAGC CAATATTCTC 240
25 AAGTTCCAAG ATACCGATGG CAAACACCTG CTGCCAAAGA TCARGGACAG CGCGGGGCAG 300
AAGGGCACAG GGAAGTGGAC CGCCATCTTC GCCCTGGGAT TACGGGGTAC CCGTCACCCT 360
CATTGGGGAA GGTGTCTTTG STCGGTGCTT ATCATCTCTT GAAGGATGAG AGAATTTCAA 420
30 GCTTGCAAAA AAGTTGAGGG GTCCCCAGAA 450

(2) INFORMATION FOR SEQ ID NO:9:

- 35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8201 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTAAAAATAC CATTAGTAA TAGTATTAGC TTTGTATTTC TGAGATTCAA CAGCAGCAGT 60
CACTTCCCTC CACTCCTATG TGTATCCCAG GACCACCCTG GGCGGGGAGG GCTGAGGTCA 120
50 GGGAGGTCTG AAGCTGGTCC TGGGCTCCGG GGGTGACAGT GATGAGGAAC TGGGTGCACA 180
CATGAGTGGG GCAGCCGGGC CTGGCCAGAG AACCAACACA CACGTGCACA GACATGTTTA 240
55 TCCACATACA CATGTGCACG CATGTGCACA AACACATTGC AGGCAGGCAT GTTGACGCCT 300
CAGGCAGCGG AGGACCCTGA CTCTGGGCCG TGCTGACCCG GGCAAGGCCC ATTGTGATGC 360

	GTGCCATGAC CTCAGAATGT CACTGGTGCT TAGCACCTAT CCGCTCTCCA GACTGCGTCT	420
	GTGTTCTACG GCAGTTACAC ACACGCAGTG GTATTCACAA GCGGTTTTGT GGACTCAAAG	480
5	GTTTTCTCCC TGAGAGGCAT AACCCAGGCC AGCTGATTCA TCAGAATCAG GTGAGTGTGA	540
	CCTGCTCTCT TCCCTCCAGG CTGACTTGGG GACAGTGGCT ATGGTATGGG CGGTGTGGC	600
10	CTCTGGGCAG CTACAGAGGA GGGTCATCCC TGAGCACTCA CCGGGCGCCC GTTCTACACT	660
	GCCCATGTAG ACGATTTTCT CTTTCGTCTT CATGGTGGCT TCGTAGAGTG GGTGCTGTTC	720
	CCAAATGTAC CCATTGACA GGTGAGCCGT CTGGGGTCAG AGAGGCAGTA ACTGGCCTGG	780
15	GAATCCAGAC AAGACCCTGG GTTTTGCTCT CAGCCCTGCT GTGTGCCATG CTAGACTTCA	840
	GGCCTCAACC CTGAGACCTC CCTGCTCTAG ATCCCAAATC TGCCAGATT TCCGATCCAA	900
20	TGGGCAGAGC CTGGCCCTGG CAGAGACACT GGGATGGATC CACTGTGGGT GGGGAGGAGG	960
	GAAGGGTCCT CAGAACACAC CTGGGGCCTA AGCTGGGTCT TGATGGTCAC TGTGGGACCC	1020
	ACTGGACACA CACAGTCCCT TGTCTGGGAG TGGCATGGGG AGCCTTCTGC CCTTGGGCAG	1080
25	TTGTGGAAAG TGAAGGAGCC CTGGAGAGCT GGCTGAGGGG AGACTATCTT CCCTTGTGTT	1140
	CAAAGGGGTC CAGGCACTGG GGCTCTCCCC AAGTATTTCT TATTCTGTCT GGCCTCGCTT	1200
30	TCCTTTTGCC CTGAGTATTC TCAGGAGGGA CGGTCCATCT AGATGTCCTC CAGGAGCAAG	1260
	GACCCACTGT TCTTCATCAG TGACCCAGGA AAATGAAGCC CCCTCCTGTG GGGACAGCTC	1320
	AGAATGGTGG AGTCCACAGT CCCTCCCTGA GAGACATGGT TTCCATGAGC ACAGTGGCTG	1380
35	CTTTGGAGAC AGTAATCATT TTCATCCCCA AAACCAAACA CACTCCTGCT CAAATGGTGT	1440
	TATTGCTAAA GCAGCTTCAC TGGTTAGACT GAAGGGCCAT GGTAGCCCAA GTGATGAGCG	1500
40	GGGTAGAATG GAGCAGTCAG GAGAGATCTT GTTCCCCGTA GGAAACTGGG CATCTCTGTG	1560
	GCCCTGAACA TCCCAGGAGG CCGATCGTAC AGAGACCTCT GGTGCCTGAC CGCAGTTCAC	1620
	ATCCACATCC CTGGAATAGA CCATCACAGG CTCTTCACCC TTGGCAGGTG GACACCATTC	1680
45	AACCTGCCGG GGCAGGATGG ACATGGTAGA GAATGCAGAT AGTTTGAGG CACAGGAGCG	1740
	GAAGGACATA CTTATGAAGT ATGACAAGGG ACACCGAGCT GGGCTGCCAG AGGACAAGGG	1800
50	GCCTGAGCCC GTTGAATCA ACAGCAGCAT TGATCGTTTT GGCATTTTGC ATGAGACGGA	1860
	GCTGCCTCCT GTGACTGCAC GGGAGGCGAA GAAAATTCGG CGGGAGATGA CACGAACGAG	1920
	CAAGTGGATG GAAATGCTGG GAGAATGGGA GACATATAAG CACAGTAGCA AACTCATAGA	1980
55	TCGAGTGATC AAGGGAATTC CCATGAACAT CCGGGGCCCG GTGTGGTCAG TCCTCCTGAA	2040
	CATTGAGGAA ATCAAGTTGA AAAACCCCGG AAGATACCAG ATCATGAAGG AGAGGGGCAA	2100

	GAGGTCATCT	GAACACATCC	ACCACATCGA	CCTGGACGTG	AGGACGACTC	TCCGGAACCA	2160
	TGTCTTCTTT	AGGGATCGAT	ATGGAGCCAA	GCAGAGGGAA	CTATTCTACA	TCCTCCTGGC	2220
5	CTATTCTGGAG	TATAACCCGG	AGGTGGGCTA	CTGCAGGGAC	CTGAGCCACA	TCACCGCCTT	2280
	GTCCTCCTT	TATCTGCCTG	AGGAGGACGC	ATTCTGGGCA	CTGGTGCAGC	TGCTGGCCAG	2340
	TGAGAGGCAC	TCCCTGCCAG	GATTCCACAG	CCCAAATGGT	GGGACAGTCC	AGGGGCTCCA	2400
10	AGACCAACAG	GAGCATGTGG	TACCCAAGTC	ACAACCCAAG	ACCATGTGGC	ATCAGGACAA	2460
	GGAAGGTCTA	TGCGGGCAGT	GTGCCTCGTT	AGGCTGCCTT	CTCCGGAACC	TGATTGACGG	2520
15	GATCTCTCTC	GGGCTCACCC	TGCGCCTGTG	GGACGTGTAT	TTGGTGGAA	GAGAACAGGT	2580
	GTTGATGCCA	ATAACCAGCA	TTGCTCTTAA	GGTTCAGCAG	AAGCGCCTCA	TGAAGACATC	2640
	CAGGTGTGGC	CTGTGGGCAC	GTCTGCGGAA	CCAATTCTTC	GATACCTGGG	CCATGAACGA	2700
20	TGACACCGTG	CTCAAGCATC	TTAGGGCCTC	TACGAAGAAA	CTAACAAGGA	AGCAAGGGGA	2760
	CCTGCCACCC	CCAGGCCCAA	CAGCCCTGGG	ACGAAGGTGT	GTGGCAGGAA	GCCCCCAGCC	2820
25	AGTCTGAACC	CTGGGGGCAG	TCCCAGGAGC	CACCCACCAT	GCCCCAACGG	CTTCCCCATG	2880
	CCAGGCAGCA	CACACCCCTC	CCTCTGGGAT	CAGCAGACTA	CAGGCGTGTC	GTCAGTGTC	2940
	GACCACAGGG	GCCACACAGA	GACCCCAAGG	ACTCCAGAGA	TGCAGCCAAA	CGCGAGCAAG	3000
30	GGTCCTTGGC	ACCCAGGCCT	GTGCCGGCTT	CACGTGGTGG	GAAGACCCTC	TGCAAGGGGT	3060
	ATAGGCAGGC	CCCTCCAGGC	CCACCAGCCC	AGTTCCAGCG	GCCCATTGTC	TCAGCTTCCC	3120
35	CGCCATGGGC	ATCTCGTTTT	TCCACGCCCT	GTCCTGGTGG	GGCTGTCCGG	GAAGACACGT	3180
	ACCTGTGGG	CACTCAGGGT	GTGCCCAGCC	TGGCCCTGGC	TCAGGGAGGA	CCTCAGGGTT	3240
	CCTGGAGATT	CCTGGAGTGG	AAGTCAATGC	CCCGGCTCCC	AACGGACCTG	GATATAGGGG	3300
40	GCCCTTGTT	CCCCATTAT	GATTTTGAAC	GGAGCTGCTG	GGTCCGTGCC	ATATCCCAGG	3360
	AGGACCAGCT	GGCCACCTGC	TGGCAGGCTG	AACACTGCCG	AGAGGTTTAC	AACRAAGATA	3420
45	TGAGTTGGCC	TGAGGAGATG	TCTTTTACAG	CAAATAGTAG	TAAAATAGAT	AGACAAAAGG	3480
	TTCCACAGA	AAAGGGAGCC	ACAGGTCTAA	GCAACCTGGG	AAACACATGC	TTCATGAACT	3540
	CAAGCATCCA	GTGCGTTAGT	AACACACAGC	CACTGACACA	GTATTTTATC	TCAGGGAGAC	3600
50	ATCTTTATGA	ACTCAACAGG	ACAAATCCCA	TTGGTATGAA	GGGGCATATG	GCTAAATGCT	3660
	ATGGTGATTT	AGTGCAGGAA	CTCTGGAGTG	GAAGTCAGAA	GAGTGTGCCC	CCATTAAAGC	3720
55	TTCGGCGGAC	CATAGCAAAA	TATGCTCCCA	AGTTTGATGG	GTTTCAGCAA	CAAGACTCCC	3780
	AAGAACTTCT	GGCTTTTCTC	TTGGATGGTC	TTCATGAAGA	TCTCAACCGA	GTCCATGAAA	3840

	AGCCATATGT	GGAAGTGAAG	GACAGTGATG	GCCGACCAGA	CTGGGAAGTA	GCTGCAGAGG	3900
	CCTGGGACAA	CCATCTAAGA	AGAAATAGAT	CAATTATTGT	GGATTTGTTC	CATGGGCAGC	3960
5	TAAGATCTCA	AGTCAAATGC	AAGACATGTG	GGCATATAAG	TGTCCGATTT	GACCCTTTCA	4020
	ATTTTTTGTG	TTTGCCACTA	CCAATGGACA	GTTACATGGA	CTTAGAAATA	ACAGTGATTA	4080
	AGTTAGATGG	TACTACCCCT	GTACGGTATG	GAATAAGACT	GAATATGGAT	GAAAAGTACA	4140
10	CAGGTTTAAA	AAAACAGCTG	AGGGATCTCT	GTGGACTTAA	TTCAGAACAA	ATCCTACTAG	4200
	CAGAAGTACA	TGATTCCAAC	ATAAAGAAGT	TTCCTCAGGA	TAACCAAAAA	GTACAACTCT	4260
15	CAGTGAGCGG	ATTTTTGTGT	GCATTTGAAA	TTCTGTCCCT	TTCATCTCCA	ATTTCACTT	4320
	CTAGTCCAAC	ACAAATAGAT	TTCTCCTCTT	CACCATCTAC	AAATGGAATG	TTCACCCTAA	4380
	CTACCAATGG	GGACCTACCC	AAACCAATAT	TCATCCCCAA	TGGAATGCCA	AACACTGTTG	4440
20	TGCCATGTGG	AACTGAGAAG	AACTTCACAA	ATGGAATGGT	TAATGGTCAC	ATGCCATCTC	4500
	TTCTGTACAG	CCCCTTTACA	GGTTACATCA	TTGCAGTCCA	CCGAAAAATG	ATGAGGACAG	4560
25	AACTGTATTT	CCTGTACACT	CAGGAGAATC	GCCCCAGCCT	CTTTGGAATG	CCATTGATTG	4620
	TTCCATGCAC	TGTGCATACC	CAGAAGAAAG	ACCTATATGA	TGCGGTTTGG	ATTCAAGTAT	4680
	CCTGGTTAGC	AAGACCACTC	CCACCTCAGG	AAGCTAGTAT	TCATGCCCAG	GATCGTGATA	4740
30	ACTGTATGGG	CTATCAATAT	CCATTCATCT	TACGAGTTGT	GCAGAAAGAT	GGGATCTCCT	4800
	GTGCTTGGTG	CCCACAGTAT	AGATTTTGCA	GAGGCTGTAA	AATTGATTGT	GGGGAAGACA	4860
35	GAGCTTTTAT	TGGAATGCC	TATATTGCTG	TGGATTGGCA	CCCCACAGCC	CTTCACCTTC	4920
	GCTATCAAAC	ATCCCAGGAA	AGGGTTGTAG	ATAAGCATGA	GAGTGTGGAG	CAGAGTCGGC	4980
	GAGCGCAAGC	CGAGCCCATC	AACCTGGACA	GCTGTCTCCG	TGCTTTTACC	AGTGAGGAAG	5040
40	AGCTAGGGGA	AAGTGAGATG	TACTACTGTT	CCAAGTGTA	GACCCACTGC	TTAGCAACAA	5100
	AGAAGCTGGA	TCTCTGGAGG	CTTCCACCCT	TCCTGATTAT	TCACCTTAAG	CGATTTCAAT	5160
45	TTGTAAATGA	TCAGTGGATA	AAATCACAGA	AAATTGTCAG	ATTTCTTCGG	GAAAGTTTTG	5220
	ATCCGAGTGC	TTTTTTGGTA	CCACGAGACC	CGGCCCTCTG	CCAGCATAAA	CCACTCACAC	5280
	CCCAGGGGGA	TGAGCTCTCC	AAGCCCAGGA	TTCTGGCAAG	AGAGGTGAAG	AAAGTGGATG	5340
50	CGCAGAGTTC	GGCTGGAAAA	GAGGACATGC	TCCTAAGCAA	AAGCCCATCT	TCACTCAGCG	5400
	CTAACATCAG	CAGCAGCCCA	AAAGGTTCTC	CTTCTTCATC	AAGAAAAAGT	GGAACCAGCT	5460
55	GTCCCTCCAG	CAAAAACAGC	AGCCCTAATA	GCAGCCCACG	GACTTTGGGG	AGGAGCAAAG	5520
	GGAGGCTCCG	GCTGCCCCAG	ATTGGCAGCA	AAAATAAGCC	GTCAAGTAGT	AAGAAGAACT	5580

	TGGATGCCAG CAAAGAGAAT GGGGCTGGGC AGATCTGTGA GCTGGCTGAC GCCTTGAGCC	5640
	GAGGGCATAT GCGGGGGGGC AGCCAACCAG AGCTGGTCAC TCCTCAGGAC CATGAGGTAG	5700
5	CTTTGGCCAA TGGATTCCTT TATGAGCATG AAGCATGTGG CAATGGCTGT GGCGATGGCT	5760
	ACAGCAATGG TCAGCTTGA AACCACAGTG AAGAAGACAG CACTGATGAC CAAAGAGAAG	5820
10	ACACTCATAT TAAGCCTATT TATAATCTAT ATGCAATTTC ATGCCATTCA GGAATTCTGA	5880
	GTGGGGGCCA TTACATCACT TATGCCAAAA ACCCAAAGTG CAAGTGGTAC TGTTATAATG	5940
	ACAGCAGCTG TGAGGAACTT CACCCTGATG AAATTGACAC CGACTCTGCC TACATTCTTT	6000
15	TCTATGAGCA GCAGGGGATA GACTACGCAC AATTTCTGCC AAAGATTGAT GGCAAAAAGA	6060
	TGGCAGACAC AAGCAGTACG GATGAAGACT CTGAGTCTGA TTACGAAAAG TACTCTATGT	6120
20	TACAGTAAAG CTACCACTCT GGCTGCTAGA CAGCTTGGTG GCGAGGGAGA TGACTCCTTG	6180
	TAGCTGATAC TTGGCAAAAG TGCTACTGAA AGACAAGCTA AATGTAGTTA TTTTATCCTG	6240
	TTAGAACAAA AATTCTAATT AAAATAGTTA ACTTGAAGAG TAGAAACAAT TGTATTTTGA	6300
25	AGTCTCATAC AAGCTGTCTG ATAGAGAACT TTCAGGCAGA TCCCACCATT AGCCTGTAAA	6360
	CAAAAGGTGT GGCACCAGCC ACCTGGGACC AAATAAGAAT TGAATTGTGC TTGTCCAGAT	6420
30	ATGAACAAAT ATGTAGTGAG TATAGAGTTT ACCAATAATC ATAACAAATA TTAAAGATTT	6480
	CCTTGGAGTC AGAGGAAAAA ACAACAATT ATAATGTTGT CTAGGGACGA CATGATACGC	6540
	TACCTCCTTT TTCCTGAAGT TTTATTCCAT TATATTGACA AGATGGAGAA AGCAAGATCA	6600
35	TGAAGGTGTG CAAATGATTC TTACGGCATG GACAAGGATT TTTCAATTTA TTTTTTAAAC	6660
	TGTTTCCATA CCCTTTCTTT TTCTTGCTTT TTGTTTTTGC CATGTGTTTT ACGTTTGAGA	6720
40	CACAACCACT CATTTGGTGGC AGGGGCATAG AGTGGTCAGT CTGAAAGGGA GGCTCTCTTA	6780
	AGAGCTATGT GCCTTCCAAC CAGAGGGAGA CCCAGTAGAA AGAAAAACAT CCTGGGAAAT	6840
	CCAGCTACCA GGGCCCTCCC AGTGGAGGCA TCTTACATTT AGGCTACTTC AAGTATCCTC	6900
45	AGAAATGTAT TCTGCACCCC CGGCCCCGCC CATGCTGAGG GAAGGGGAGC AGTTGCCAAT	6960
	ATTTGCACCA TCTTCACATG CACATGTTGC AACAAGAGCT TCTGGGAAGG TAAGCGGCAT	7020
50	CGGAGCTAGA TCACGTTTCA CAATTAGTGG TTATTCTTTT CTGTGTTTGT TTTGCACTTT	7080
	AAAAAAGAGA GAACACATGC AAATGAACTT GCTTGTGTGT ATTTGATGGC TCTAAGGGCT	7140
	ATAAATTACA AACAAAACAC ATCCCAGACA TTAGGAGTTC ATAAGTATAT TTAATGAAAT	7200
55	TGGTGGTTTT AGGAAGTCAA CTTTAGTTTT GCTTTGTTTG CATGTCCACT GGTTTTTTTA	7260
	TTTTGATATT TGTCTTTTTT TAAATTTTAC AGTAGTCATT GAAAGTTATG TTTCTTTGCT	7320

TACTTCATTT TTTCCCTCTA ATTATTTAAG ATTGGAACAA AAGTATAAAT ATTATTTATT 7380
TGAGGTAGAA TTTTTTTCAT GTAGTTTCTT AATATATACT TGAAGGAAAT GTTTCACCTT 7440
5 ATTTTGGTC TTTGTTTATT CATTTAGACC CTGCAAGTTG ATTCTCATTG CCAGATTCCA 7500
TTACCCTTTC TTCCTCATAG GTAGTAATTA CCAATGTAAC TAAGCATTTG TGTTCTGATA 7560
TCTGAGGCCA GTAACATTA ATATCTAGTT CTCAGAGCAT TTGGAAAGGT TATCTTAAAT 7620
10 GGCTACCTAA ATTGAAATCC TTTTCAGAAA AAATATAATT GCAAGTAGGT AGGAGTGGCC 7680
TAAATTGTCT AATGTAATAA AGTCAGACAA AATGCACACT TTATAGTTTC AAGATTTTCA 7740
15 GTAAATAAAA TCTGTCCATT CCTACCTGGA CATGTCCCAT TAAAAAGTGG AAGATTTTAA 7800
ATAATTTCTT TACAGATGTT TTATTTAAAC AGGTAGCACA ATCTACTAAT GTTGTGTGAT 7860
TTGTGTTATA CTGGTTGTAA TTAATTTTTT TAATTCATGA ACTAGCGGAA AATTTATTAA 7920
20 ATTAATATT AACTACATTC ACCTGTAAA TTACTGTATA AACTTGTTG ACAATGCACT 7980
GACTTTAGAA AGATGTTAAT GTACATAAAT AGAGTGTAAT TAAAATAGTG TTGATGTACT 8040
25 GAAATATGAA CTGTATCAAA AGTATTGGTA ATTGTATATG GGGTGACCT GTTTATCTGT 8100
TAACTATTAT CCAAACAAAT TAAATACTGT GGTGCCTCT ATGTGCTGTT TTTCCTCATA 8160
CAAGTAAACA CAGAAAGTCA AAAAAAAAAA AAAAAAAAAA A 8201
30

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 945 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

45 GAATTCTGCG GCCGCCAGAA AATTCACAAA GAGATGCCCT GTAAGTGTA TGTATGTGGC 60
AGTGACTTCT GCCATACTTC ATACCTACTT GAACATCAGA GGGTCCATCA TGAAGAGAAA 120
GCCTATGAGT ATGATGAATA TGGGTTGGCC TATATTAAAC AACAAGGAAT TCATTTTACA 180
50 GAAAAGCCCT ATACGTGTAG TGAATGTGGA AAAGACTTCA GATTGAATTC ACATCTTATT 240
CAGCATCAAA GAATTCACAC AGGAGAGAAA GCACATGAAT GTCATGAATG TGGAAAAGCT 300
TTCAGTCAAA CCTCATGCCT TATTCAGCAT CACAAAATGC ATAGGAAAGA GACTCGTATT 360
55 GAATGTAATG AGTATTGAGG GCAGGTTCAA GTCATAGCTC AGATCTTATC CTGCAACAAG 420

GAAGTCCTCA CCAGACAGAA AGCCTTTGAT TGGTGATGTA TGGGAAAAGA ACTCCAGTCA 480
 GAGAGCACAT CTAGTTCAAC ATCAGAGCAT TCATACCAA GAGAACTCAT GAATGTAATG 540
 5 AAGATGGGAA GATATTTATC AAATTCAGGC TTCATTCAGC ATCTGAGAGT TCACACCAGG 600
 GAGCAAATCA TGTATGTACT GCATGTGGTA AAGCCTTCAG TCATAGCTCA GCCATTGCTC 660
 AGCATCAGAT AATTCACACC AGAGAGAAAC CCTCTGAATG TGACGAATGA AGAAAAGGTA 720
 10 TTAGTGTTAA ACTCTTAATC GACTCCTGCA AATCTATACC AGTGAGAAAT CTTACAAATG 780
 TATTGGATTG TGGCAAATTT CTCATGCTAT TAGTATTTTC ATACCTTAGT CACATGTGGG 840
 15 GGAATCCACA TGGGAATAAA CTCCCATTCG TGCAATGATT GTGAAAAGCA TCAGGCAAGG 900
 AACTTCCTGG TTAGGTTCAA TTCCACGCCA TGCAAAAGGT TTTTA 945

(2) INFORMATION FOR SEQ ID NO:11:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 971 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

25

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAATTCTGCG GCCGCCTCTT CGCTGAGGCG GGGCCAGACT TTGAACTGCG GTTAGAGCTG 60
 35 TATGGGGCCT GTGTGGAAGA AGAGGGGGCC CTGACTGGCG GCCCCAAGAG GCTTGCCACC 120
 AAATCAGCA GCTCCCTGGG CCGCTCCTCA GGGAGGCGTG TCCGGGCATC GCTGGACAGT 180
 GCTGGGGGTT CAGGGAGCAG TCCCATCTTG CTCCCCACCC CAGTTGTTGG TGGTCTCGT 240
 40 TACCACCTCT TGGCTCACAC CAACTCACC CTGGGAGGAG TGCAAGATGG ATTCCGCACA 300
 CATGACCTCA CCCTTGGCAG TCATGAGGAG AACCTGCCTG GCTGCCCCCTT TATGGTAGCG 360
 45 TGTGTTGCCG TCTGGCAGCT CAGCCTCTCT GCATGACTCA GCCCACTGCA AGTGGTACCC 420
 TCAGGGTGCA GCAAGCTGGG GAGATGCAGA ACTGGGCACA AGTGCATGGA GTTCTGAAAG 480
 GCACAAACCT CTTCTGTTAC CGGCAACCTG AGGATGCAGA CACTGGGGAA GAGCCGCTGC 540
 50 TTACTATTGC TGTCAACAAG GAGACTCGAG TCCGGGCAGG GGAGCTGGAC CAGGCTCTAG 600
 GACGGCCCTT CACCCTAAGC ATCAGTAACC AGTATGGGGA TGATGAGGTG ACACACACCC 660
 55 TTCAGACAGA AAGTCGGGAA GCACTGCAGA GCTGGATGGA GGCTCTTGTT GCAGCTTTTT 720
 CTTTTGGACA ATGAGCCAAT GGAAGCAGTG CTTGTGATGA AATCAATGAA AATTGGAAAC 780

TTCTGCTCC CCCGAAACC ACCCAAGCA CTGGCAAAGC AGGGGGTCCT TGTACCATGA 840
GATGGCTATT GAGCCGCTGG ATGACATCGC AGCGGGTGAA AGACATCCTG ACCCAGGGGG 900
5 AGGGCGCAAG GTTGGAGACA CCCCCCCGG TTGGAATTTT TACAGACAGC CTGCCTGCTT 960
ACCCCTGTCTG C 971

(2) INFORMATION FOR SEQ ID NO:12:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1285 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAATTCGGCA CGAGAGCAAG CAAGAGAAAG AGAAGAGCAA GAAGAAAAAA GGAGGTAAAA 60
25 CAGAACAGGA TGGCTATCAG AAACCCACCA ACAAACACTT CACGCAGAGT CCCAAGGAAG 120
TCAGTGGCCG ACCTGCTGGG GTCCTTTGGA AGGCAAACGA AGGACTCCTT CTGATCACTG 180
CTCCCAAGGC TGAGGAACAA CAACGTGATG AATATCTGGA AAGTTTCTGC AAGATGGCTA 240
30 CCAGGAAAAT CTCTGTGATC ACCATCTTCG GCCCTGTCAA CAACAGCACC ATGAAAATCG 300
ACCACTTTCA GCTAGATAAT GAGAAGCCCA TCGGAGTGGT GGATGATGAA GACTTGGTAG 360
35 ACCAGCGTCT CATCAGCGAG CTGAGGAAAG AGTACGGAAT GACCTACAAT GACTTCTTCA 420
TGGTGCTAAC AGATGTGGAT CTGAGAGTCA AGCAATACTA TGAGGTACCA ATAACAATGA 480
AGTCTGTGTT GGATCTGATC GATACTTTCC AGTCCCGAAT CAAAGATATG GAGAAGCAGA 540
40 AGAAGGAGGG CATTGTTTGC AAAGAGGACA AAAAGCAGTC CCTGGAGAAC TTCCTATCCA 600
GGTTCCGGTG GAGGAGGAGG TTGCTGGTGA TCTCTGCTCC TAACGATGAA GACTGGGCCT 660
45 ATTCACAGCA GCTCTCTGCC CTCAGTGGTC AGGCGTGCAA TTTTGGTCTG CGCCACATAA 720
CCATTCTGAA GCTTTTAGGC GTTGGAGAGG AAGTTGGGGG AGTGTTAGAA CTGTTCCCAA 780
TTAATGGGAG CTCTGTTGTT GAGCGAGAAG ACGTACCAGC CCATTGGGT GAAAGACATC 840
50 CGTAACTATT TCAAGTGAGC CCGGAGTACT TCTCCATGCT TCTAGTCGGA AAAGACGGAA 900
ATGTCAAATC CTGGTATCCT TCCCCAATGT GGTCCATGGT GATTGTGTAC GATTTAATTG 960
55 ATTCGATGCA ACTTCGGAGA CAGGAAATGG CGATTCAGCA GTCAGTGGG ATGCGCTGCC 1020
CAGAAGATGA GTATGCAGGC TATGGTTACC ATAGTTACCA CCAAGGATAC CAGGATGGTT 1080

ACCAGGATGA CTACCGTCAT CATGAGAGTT ATCACCATGG ATACCCTTAC TGAGCAGAAA 1140
 TATGTAACCT TAGACTCAGC CAGTTTCCTC TGCAGCTGCT AAAACTACAT GTGGCCAGCT 1200
 5 CCATTCTTCC ACACTGCGTA CTACATTTC TGCCTTTTTC TTTCAGTGTT TTTCTAAGAC 1260
 TAAATAAATA GCCAACTTTC ACCTT 1285

10 (2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1439 base pairs
 (B) TYPE: nucleic acid
 15 (C) STRANDEDNESS: both
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAATTCTGCG GCCGCCATTA CTCCTGCAAC ATATCTGGCT CTCTGAAGCG GCACTACAAC 60
 25 AGGAAGCACC CTAATGAGGA GTATGCCAAC GTGGGCACCG GGGAGCTGGC AGCGGAGGTG 120
 CTCATCCAGC AAGGTGGTTT GAAGTGTCTT GTTTCAGCT TTGTATATGG CACCAAATGG 180
 30 GAGTTCAATA GGCACCTGAA GAACAAACAT GGCTTGAAGG TGGTGGAAAT TGATGGAGAC 240
 CCCAAGTGGG AGACAGCAAC AGAAGCTCCT GAGGAGCCCT CCACCCAGTA TCTCCACATC 300
 ACAGAGTCCG AAGAAGACGT TCAAGGGACA CAGGCAGCGG TGGCCGCGCT CCAGGACCTG 360
 35 AGATACACCT CTGAGAGTGG CGACCGACTG GACCCACCG CTGTGAACAT CCTGCAGCAG 420
 ATCATTGAGC TGGGCGCCGA GACCATGAC GCCACTGCCC TTGCCTCGGT GGTGCCATG 480
 40 GCACCAGGGA CGGTGACTGT GGTTAAGCAG GTCACCGAGG AGGAGCCCAG CTCCAACCAC 540
 ACGGTCATGA TCCAGGAGAC GGTCCAGCAA GCGTCCGTGG AGCTTGCCGA GCAGCACCAC 600
 CTGGTGGTGT CCTCCGACGA CGTGGAGGGC ATTGAGACGG TGACTGTCTA CACGCAGGGC 660
 45 GGGGAGGCCT CGGAGTTTCA CTCTACGTG CAGGAGGCCA TGCAGCCTGT GGAGGAGCAG 720
 GCCTGTGGAG CAGCCGGCCC AGGAACTCTA GAGGACATGT GGCATCGGAT GGCCACAGGG 780
 50 CGGGGCTGTC CAGGCTCTTC AGGCACCCAG GGTGGGGAGG CCACCTTCCT GCCCTACCCG 840
 AGAATGGTGT CTCCTTTGCC CTCCTGCCC AGCAGCCTGA TAGGACTCTC CTAGTCCAAC 900
 TTGGGGTGGG CAAGGCAGTC AGCATACCA GCAACACCAC AGGACCCTCA CCCCAGCATA 960
 55 GACACACACC CCCTGACCCT TACCATCTGC TTCCTGAAAG ACTTCAGTGT CAGCTCCCCT 1020
 ACACACACCC CACACCTTCA CCCCTTGCTT CAAGATTCAA ACAGAGACTC CCACTCCCCC 1080

TCAGCATCTT CCCTGGATCA CAACCCCAGC TCCTTGACCC CCATCTAGGT GCCAAATGTT 1140
CATCTGCAAC CGCTATGCAG TCTGGTGAGA GGGAGACAGC CATCACATAG AAAGTGGCCG 1200
5 TACGGGTTTT TAATCACTGC TGGGTGGGGT GGGGGTAGGG GGATTGTCCT GGCTTTGTCG 1260
ACAAAGTCCC ACTTCCCCGA GTATTAAGGG CCCTTGATAT CAAGTGAGGT AAATTCACCC 1320
10 ATCACAGGGT CTCGCCCTAC CATCCTGGAA TTATTTCACT TTTAAGATAA ATGCACTATT 1380
TCACTGTTCG CCTCCCATTC TAAGGAGGTG AGGTGGTTGG AATAAAAACA GTTCCTGTC 1439

15 (2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 349 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAATTCTGCG GNCGCGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 60
30 AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 120
AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 180
35 GGTCTTAACA CATATGGGAC TGATGTCATC TCGACCTCTC CATTATTGA GTCTGTGATT 240
TATTTGGAGT GGAGGCATCG TTTTAAAGAA ACACATGTCA TCTAGGTTGT CTAACCTAT 300
CTGCATCTAC TCTCACCTCA NCCCCCCCCC CCCCTTCCCC CCCTNTTCC 349
40

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 572 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

55 GAATTCTGCG GCCGCCGATC CGAGGTCCTT TTAGTCTCAG AGGATGGGAA GATCCTGGCA 60
GAAGCAGATG GACTGAGCAC AAACCACTGG CTGATCGGGA CAGACAAGTG TGTGGAGAGG 120

ATCAATGAGA TGGTGAACAG GGCCAAACGG AAAGCAGGGG TGGATCCTCT GGTACCGCTG 180
CGAAGCTTGG GCCTATCTCT GAGCGGTGGG GACCAGGAGG ACGCGGGGAG GATCCTGATC 240
5 GAGGAGCTGA GGGACCGATT TCCCTACCTG AGTGAAAGCT ACTTAATCAC ACCGACGGCG 300
GCGGCTCCAT CGACACAGCT ACACCGGATG GTGGAGTTGT GCTCATATCT GGAACAGGCT 360
CCAAGTGCAG GCTCATCAAC CCTGATGGCT CCGAGAGTGG CTGCGGGCGG CTTGGGGGCA 420
10 TATTATGGGT GATGAGGGTT CAGCCTACTG GATCGCACAC CAAGCAGTGA AAATAGTGTT 480
TGGACTCCAT TGAAACTAG AGGCGGTCCC ATGATATCGG TTACGTCAA CAGGCCATGT 540
15 TCCACTATTT CCAGGTTT CAG ATCCGCTAGG TT 572

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
20 (A) LENGTH: 402 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

30 GAATTCTGCG GCCGCCAGAG CAGCACGGAG ATCAGCAAGA CGCGGGGCGG GGAGACAAAG 60
CGCGAGGTGC GGGTGGAGGA GTCCACCCAG GTCGGCGGGG CACCCCTTCC CTGCTGTGTT 120
35 TGGGGACTTC CTGGGCCGGG AGCGCCTGGC ATCCTTCGGC AGTATCACCC GGCAGCAGGA 180
GGGTGAGGCC AGCTCTCAGG ACATGACTGC ACAGGTGACC AGCCCATCGG GCAAGGTGGA 240
AGCCGCAGAG ATCGTCGAGG GCGAGGACAG CGTCTACAGC GTGCGCTTTG TGCCCCAGGA 300
40 AATGGGGGCC CATACGGTCG GTGTCAAGTA CCGTGNCAG CACGTGCCCC GNAGNCCCTT 360
TCAGTTCACT GTNGGGCCGC TGGGTGANGG TTGGTGCCCA CA 402

45 (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 771 base pairs
(B) TYPE: nucleic acid
50 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AAGGGGAAGA GAAGAGAGTG TCCAGGGAGC CAGCAGGTGT CCTCTCCCAG AGTGGTATGC 60
AGCTGGAATA TCTGTCCCTC CCCTTCCAAC TTCCCGCACG CAGATCCTTG CAGGTTGAGC 120
5 TCTGTGGAGG CCAACCTGTC CTCTCCAGGG TGAAAGTGCA GTGGAGGCCT TCTGGCTCCA 180
CTCCAAATGT GATAGAAGGG GATCTCCTGG TATTTGGCCA GCAGCTTGCT CCTCCAATGG 240
GCATGGGGGA GGTGATGGAG GAAGAGCGCA GGTGTGTGTA ACTGTCCTTG AACATTAGCG 300
10 GTTTCGGCTC CTCCACCAAG TATCCGCCCA GAGTCCGCTC CAGCTCCAGC ACCTCCTTCA 360
GTGCTACAGG CCTGTCTCC AGACAGTAGA CCCGGAGTCT GTACTCCAGG GAGGTGCAGA 420
15 GGGCGGGGGC GAAGACGGCC AGCTGGASCC GCTTGACTGC TGAGCGGGAA TAGGACTCGC 480
CCGTGAACAC GTAGGTGCCC AGCTGGTCCA GCAGGATGTG ACAGGCCCTG GGCTCCAGCT 540
GGCAGTAGCA GGGTGTGTC AGGGTCTCCT CATCCAGGGT CACCACCTCC TCCCAGTGGC 600
20 CCTGGTGGGC CTGGGTCTTG AGCTGAAAGA TCCAGTCACG GGCAGTACT TCGGCACAGT 660
GGGGCATGGT GAGGATGACG GGGCGGCACA GCAGGAGGCC TGTGGGTCCA CAGGTCACCG 720
25 AGGGGCTCAA TACTGTCTCG GGAGAGGCAT AATCTGGCAC ATCATAAGGG T 771

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
30 (A) LENGTH: 638 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: cDNA

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAATTCTGCG GNCGCGCCCT ACATGTGAAC AACGATCGGG CAAAAGTGAT CCTGAAGCCA 60
GACAAGACTA CTATTACAGA ACCACACCAC ATCTGGCCCA CTCTGACTGA CGAAGAATGG 120
45 ATCAAGGTCG AGGTGCAGCT CAAGGATCTG ATCTTGCTG ACTACGGCAA GAAAAACAAT 180
GTGAACGTGG CATCACTGAC ACAATCAGAA ATTCGAGACA TCATCCTGGG TATTGAGGAT 240
CTTCGGGAAC CGTCACAGGA GGGGGAGNAG ATCGCTGAGA TCCGAGAAGC AGGCCCAGGG 300
50 AACAATCGCA GGTGACGGC AACACAGGAT TCGCACTTGT CAACAAGCAT TGGGGATGAG 360
TTCAACAACC TCCACCACCC CAGGAATTTT TGAGACCCCG GNTTTTCCTC CATCCNAGNN 420
55 TTTANTTGGG GGGGTCAAAG GGCCNNTTNT TTTTGCCAC CCTGAACCCT AGGGCCCAAC 480
CCNNTTTTTT TTCNACNTT TNGGAATNAA AGGGGNTTTG NTCANACCCC ANCCCCCCN 540

GNTTTNNTTT NGNNGGTCCC CTTTNTTTTT TTCCCCCNG NCCCNNTTG NNGGTTTCCTT 600
TTTGGGGGGC CCCCNTTCN CCCCGGGNNG GGGCCCCC 638

5 (2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 2056 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- 20 (A) NAME/KEY: -
(B) LOCATION: 176
(D) OTHER INFORMATION: /label= ATG
/note= "start codon"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

25 GAATTCGGCA CGAGGTTTTT TTTTTTTTTT TTTTTTTTTT TTATATGCAT GGAGTTATAC 60
AGGATGTGAC TTTTGTAGAT TGGCTTTTTC CGTTGACTAT CCTGCCCTG AGATCCACCC 120
AAGTTGTGGG ATCTGAAACT TGCCACCCT TCGGGATATT GCAGGACGCT GCATCATGAG 180
30 CGACAGTAAA TGTGACAGTC AGTTTTATAG TGTGCAAGTG GCAGACTCAA CCTTCACTGT 240
CCTAAACGT TACCAGCAGC TGAACCAAT TGGCTCTGGG GCCCAAGGGA TTGTTTGTGC 300
35 TGCATTTGAT ACAGTTCTTG GGATAAATGT TGCAGTCAAG AAATAAGCC GTCCTTTTCA 360
GAACCRAACT CATGCAAAGA GAGCTTATCG TGAAGTTGTC CTCTTAAAT GTGTCAATCA 420
TAAAAATATA ATTAGTTTGT TAAATGTGTT TACACCACAA AAACTCTAG AAGAATTTCA 480
40 AGATGTGTAT TTGGTTATGG AATTAATGGA TGCTAACTTA TGTGAGTTA TTCACATGGA 540
GCTGGATCAT GAAAGAATGT CCTACCTTCT TTACCAGATG CTTTGTGGTA TTAACATCT 600
45 GCATTCAGCT GGTATAATTC ATAGAGATTT GAAGCCTAGC AACATTGTTG TGAAATCAGA 660
CTGCACCCTG AAGATCCTTG ACTTTGGCCT GGCCCGGACA GCGTGCACTA ACTTCATGAT 720
GACCCCTTAC GTGGTGACAC GGTACTACCG GGCGCCGAA GTCATCCTGG GTATGGGCTA 780
50 CAAAGAGAAC GTGGATATCT GGTCACTGGG TTGCATCATG GGAGAGCTGG TGAAAGGTTG 840
TGTGATATTC CAAGGCACTG ACCATATTGA TCAGTGAAT AAAGTTATTG AGCAGCTGGG 900
55 AACACCATCA GCAGAGTTCA TGAAGAACT TCAGCCAAT GTGAGGAAT ATGTGAAAA 960
CAGACCAAAG TTTCCTGGAA TCAAATTGGA AGAAGTCTTT CCAGATTGGT TATTCCTATC 1020

AGAATCTGAG CGAGACAAAA TAAAAACAAG TCAAGCCAGA GATCTGTTAT CACAAATGTT 1080
AGTGATTGAT CCTGACAAGC GGATCTCTGT AGACGAAGCT CTGCGTCACC CATACATCAC 1140
5 TGTTTGGTAT GACCCCGCCG AAGCAGAAGC CCCACCACCT CCAATTTATG ATGCCCAGTT 1200
GGAAGAAAGA GAACATGCAA TTGAGGAATG GAAAGAGCTA ATTTACAAAG AAGTCATGGA 1260
TTGGGAAGAA AGAAGCAAGA ATGGTGTGT AAAAGATCAG CCTTCAGCAC AGATGCAGCA 1320
10 GTAAGTAGCA ACGCCACTCC TTCTCAGTCT TCATCGATCA ATGACATTTT ATCCATGTCC 1380
ACTGAGCAGA CGCTGGCCTC AGACACAGAC AGCAGTCTTG ATGCCTCGAC GGGACCCCTT 1440
15 GAAGGCTGTC GATGATAGGT TAGAAATAGC AAACCTGTCA GCATTGAAGG AACTCTCACC 1500
TCCGTGGGCC TGAAATGCTT GGGAGTTGAT GGAACCAAAT AGAAAACTC CATGTTCTGC 1560
ATGTAAGAAA CACAATGCCT TGCCCTACTC AGACCTGATA GGATTGCCTG CTTAGATGAT 1620
20 AAAATGAGGC AGAATATGTC TGAAGGAAAA AATTCCAACC ACACTTCTAG AGATTTTGTC 1680
CAAGATCATT TCAGGTGAGC AGTTAGAGTA GGTGAATTTG TTTCCAAATT GTACTAGTGA 1740
25 CAGTTTCTCA TCATCTGTAA CTGTTGAGAT GTATGTGCAT GTGACCACCA ATGCTTGCTT 1800
GGACTTGCCC ATCTAGCACT TTGGGAATCA GTATTTAAAT GCCCAATAAT CTTCCAGGTA 1860
GTGCTGCTTC TGGAGTTATC TCCTAATCCT CCTAAGTAAT TTGGTGTCTG TCCAGGAAAA 1920
30 GTCGATTAT GTGTATTAAT TGGCCATCAT GATGTTATCA TATCTTATTC CCCTTTATGC 1980
TATGATTAT TCTATCTTTT GTATTTTCTG AGACATATAA TTAAATCTAT TTAATAAATA 2040
35 AAAATATATA GCTTTT 2056

(2) INFORMATION FOR SEQ ID NO:20:

- 40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
45 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAATTCTGCG GTCGCCACGA AGAGAACATG CATGATCTTC AGTACCATAC CCACTACGCC 60
CAGAACCGCA CTGTGGAGAG GTTTGAGTCT CTGTAAGAC GCATGGCTTC TCACGAGATT 120
55 GAAATTGGCA CCATCTTCAC CAACATCAAT GCCACCGACA ACCACGCGCA CAGCATGCTC 180
ATGTACCTGG ATGACGTGCG GCTCTCCTGC ACGCTGGGCT TCCACACCCA TGCCGAGGAG 240

CTCTACTACC TGAACAAGTC TGTCTCCATC ATGCTGGGCA CCACAGACCT GCTCCGGGAG 300
CGCTTCAGCC TGCTCAGTGC CCGGCTGGAC CTCAACGTCC GGAACCTCTC CATGATCGTG 360
5 GAGGAGATGA AGGGAGGGGA CACACAGAAT GGGGAGATCC TTCGGAATGT AACATCCTAC 420
GAGGTGCCCC CGGCCTCCAG GACCAAGAGG TTCAAAAGAG ATTTGGCGTG AAACGGCTGT 480
10 GGCGGAGAGG CCAAAGGAGA CCG 503

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 1618 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear
20 (ii) MOLECULE TYPE: cDNA
(ix) FEATURE:
25 (A) NAME/KEY: -
(B) LOCATION: 58
(D) OTHER INFORMATION: /label= atg
/note= "start codon"

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
GAATTCTGCG GCCGCCGCCG CCACCCGAGC CGGAGCGGGT TGGGCCGCCA AGGCAAGATG 60
GTGGACTACA GCGTGTGGGA CCACATTGAG GTGTCTGATG ATGAAGACGA GACGCACCCC 120
35 AACATCGACA CGGCCAGTCT CTTCCGCTGG CGGCATCAGG CCCGGGTGGA ACGCATGGAG 180
CAGTTCCAGA AGGAGAAGGA GGAAGTGGAC AGGGGCTGCC GCGAGTGCAA GCGCAAGGTG 240
40 GCCGAGTGCC AGAGGAAACT GAAGGAGCTG GAGGTGGCCG AGGGCGGCAA GGCAGAGCTG 300
GAGCGCCTGC AGGCCGAGAG CACAGCAGCT GCGCAAGGAG GAGCGGAGCT GGGAGCAGAA 360
GCTGGAGGGA GATGCGCAAG AAGGAGAAGA GCATGCCCTG GCAACGTGGA CACGCTCAGC 420
45 AAAGACGGCT TCAGCAAGAG CATGGTAAAT ACCAAGCCCG AGAAGACGGA GGAGGACTCA 480
GAGGAGGTGA GGGAGCAGAA ACACAAGACC TTCGTGGAAA AATACGAGAA ACAGATCAAG 540
50 CACTTTGGCA TGCTTCGCCG CTGGGATGAC AGCCACAAGT ACCTGTCAGA CAACGTCCAC 600
CTGGTGTGCG AGGAGACAGC CAATTACCTG GTCATTTGGT GCATTGACCT AGAGGTGGAG 660
GAGAAATGTG CACTCATGGA GCAGGTGGCC CACCAGACAA TCGTCATGCA ATTTATCCTG 720
55 GAGCTGGCCA AGAGCCTAAA GGTGGACCCC CGGGCCTGCT TCCGGCAGTT CTTCACTAAG 780
ATTAAGACAG CCGATCGCCA GTACATGGAG GGCTTCAACG ACGAGCTGGA AGCCTTCAAG 840

5 GAGCGTGTGC GGGGCCGTGC CAAGCTGCGC ATCGAGAAGG CCATGAAGGA GTACGAGGAG 900
GAGGAGCGCA AGAAGCGGCT CGGCCCCGGC GGCCTGGACC CCGTCGAGGT CTACGAGTCC 960
CTCCCTGAGG AACTCCAGAA GTGCTTCGAT GTGAAGGACG TGCAGATGCT GCAGGACGCC 1020
ATCAGCAAGA TGGACCCAC CGACGCAAAG TACCACATGC AGCGCTGCAT TGA CTCTGGC 1080
10 CTCTGGGTCC CCAACTCTAA GGCCAGCGAG GCCAAGGAGG GAGAGGAGGC AGGTCCTGGG 1140
GACCCATTAC TGGAAGCTGT TCCCAAGACG GGGCGATGAG AAGGATGTCA GTGTGTGACC 1200
TGCCCCAGCT ACCACCGCCA CCTGCTTCCA GGGCCCTATG TGCCCCCTT TCAAGAAAAC 1260
15 AAGATAGATG CCATCTCGCC CGCTCCTGAC TTCCTCTACT TGCCTGCTC GGGCCAGCCT 1320
GGGGGGCCG CCCAGCCCTC CTGGCCTCT CCACTGTCTC CACTCTCCAG CGCCAATCA 1380
20 AGTCTCTGCT TTGAGTCAAG GGGCTTCACT GCCTGCAGCC CCCCATCAGC ATTATGCCAA 1440
AGGCCCGGG GTCCGGGGAA GGGCAGAGGT CACCAGGCTG GTCTACCAGG TAGTTGGGGA 1500
GGGTCCCCAA CCAAGGGGCC GGCTCTCGTC ACTGGGCTCT GTTTTCACTG TTCGTCTGCT 1560
25 GTCTGTGTCT TCTAATTGGC AAACAACAAT GATCTTCCAA TAAAAGATTT CAGATGCC 1618

30 (2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 329 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
35 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GAATTCTGCG GCCGCGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 60
45 AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG 120
AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGTCTCTATG ATCTTTCCAT TCAAACTTC 180
CAAGTTTCTC CTTATGTGGA ACCGAAATCT TTCTTTCTCC CGCGAACTT TACTACTATC 240
50 AGATAATTGA AGACAGATCT CTGTGTGTTT TCTCAAGCC CAAACCAATT CTGTTCTTC 300
ACTCTATATA GTGGTAATAT GAATGTTTA 329

55 (2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 391 base pairs

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

10 GAATTCGGCA CGAGGTTTTT TTTTTTTTTT TTTTTTTTTT TTTTGAAT GGGGTTATCC 60
 AGGATGTGAC TTTTGGAGAT TGGTTTTTTC CGTGGATTAT CCTGCCCCTG AGATCCACCC 120
 15 AAGTTGTGGG ATCTGAACT TGCCCACCCT CCGGGATTTT GAAGGACGCT GAATCATGAG 180
 CGACAGTAAT TGTGAAAGCC AGTTTTTTTG TGTGAAAGTG GAAGACTCAA CCTCCACTGT 240
 CCTAAAACGT TACCAGAAGT TGAAACCAAT TGGCTCTGGG GCCCAAGGGA TTGTCGGGGC 300
 20 TGCATCGGGT ACAGTTCTTG GGGATAAATG TTGGAGCCAA GGAATTAAGC CCGCCCCCTT 360
 TCAGAACCCA ACTCATGAAA GGGAGTTCTC C 391

25 (2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 148 amino acids
 (B) TYPE: amino acid
 30 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

40 Met Asp Thr Asp Thr Asp Thr Phe Thr Cys Gln Lys Asp Gly Arg Trp
 1 5 10 15
 Phe Pro Glu Arg Ile Ser Cys Ser Pro Lys Lys Cys Pro Leu Pro Glu
 20 25 30
 45 Asn Ile Thr His Ile Leu Val His Gly Asp Asp Phe Ser Val Asn Arg
 35 40 45
 Gln Val Ser Val Ser Cys Ala Glu Gly Tyr Thr Phe Glu Gly Val Asn
 50 55 60
 Ile Ser Val Cys Gln Leu Asp Gly Thr Trp Glu Pro Pro Phe Ser Asp
 65 70 75 80
 55 Glu Ser Cys Ser Pro Val Ser Cys Gly Lys Leu Ser Lys Val Gln Asn
 85 90 95
 Met Asp Leu Trp Leu Ala Val Asn Thr Pro Leu Xaa Ser Thr Ile Ile

100 105 110

Tyr Gln Cys Glu Pro Gly Tyr Glu Gly Gly Gly Glu Gln Gly Thr Cys
115 120 125

Leu Pro Gly Glu Gln Thr Val Glu Trp Arg Gly Gly Asn Met Gln Arg
130 135 140

Asp Gln Val Xaa
145

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 138 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Glu Leu Leu Ala Ala His Gly Thr Leu Glu Leu Gln Ala Glu Ile Leu
1 5 10 15

Pro Arg Arg Pro Pro Thr Pro Glu Ala Gln Ser Glu Glu Glu Arg Ser
20 25 30

Asp Glu Glu Pro Glu Ala Lys Glu Glu Glu Glu Lys Pro His Met
35 40 45

Pro Thr Glu Phe Asp Phe Asp Asp Glu Pro Val Thr Pro Lys Asp Ser
50 55 60

Leu Ile Asp Arg Arg Arg Thr Pro Gly Ser Ser Ala Arg Ser Gln Lys
65 70 75 80

Arg Glu Ala Arg Leu Asp Lys Val Leu Ser Asp Met Lys Arg His Lys
85 90 95

Lys Leu Glu Glu Gln Ile Leu Arg Thr Gly Arg Asp Leu Phe Ser Leu
100 105 110

Asp Ser Glu Asp Pro Ser Pro Ala Ser Pro Pro Leu Arg Ser Ser Gly
115 120 125

Ser Ser Leu Phe Pro Arg Gln Arg Lys Tyr
130 135

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 215 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

5

(v) FRAGMENT TYPE: internal

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Val Gly Thr Glu Glu Asp Gly Gly Gly Val Gly His Arg Thr Val Tyr
1 5 10 15

15

Leu Phe Asp Arg Arg Glu Lys Glu Ser Glu Leu Gly Asp Arg Pro Leu
20 25 30

Gln Val Gly Glu Arg Ser Asp Tyr Ala Gly Phe Arg Ala Cys Val Cys
35 40 45

20

Gln Thr Leu Gly Ile Ser Pro Glu Glu Lys Phe Val Ile Thr Thr Thr
50 55 60

25

Ser Arg Lys Glu Ile Thr Cys Asp Asn Phe Asp Glu Thr Val Lys Asp
65 70 75 80

Gly Val Thr Leu Tyr Leu Leu Gln Ser Val Asn Gln Leu Leu Leu Thr
85 90 95

30

Ala Thr Lys Glu Arg Ile Asp Phe Leu Pro His Tyr Asp Thr Leu Val
100 105 110

Lys Ser Gly Met Tyr Glu Tyr Tyr Ala Ser Glu Gly Gln Asn Pro Leu
115 120 125

35

Pro Phe Ala Leu Ala Glu Leu Ile Asp Asn Ser Leu Ser Ala Thr Ser
130 135 140

40

Arg Asn Ile Gly Val Arg Arg Ile Gln Ile Lys Leu Leu Phe Asp Glu
145 150 155 160

Thr Gln Gly Lys Pro Ala Val Ala Val Ile Asp Asn Gly Arg Gly Met
165 170 175

45

Thr Ser Lys Gln Leu Asn Asn Trp Ala Val Tyr Arg Leu Ser Lys Phe
180 185 190

Thr Arg Gln Gly Asp Phe Glu Ser Asp His Ser Gly Cys Ser Ser Ser
195 200 205

50

Thr Ser Ala Thr Gln Phe Lys
210 215

55

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
1 5 10 15
15 Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
20 25 30
Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Ser Ile Arg Pro Asp
35 40 45
20 Met Ser Arg Ser Val Ala Leu Asp Val Leu Ala Leu Leu Ser Leu Ser
50 55 60
25 Cys Leu Glu Ala Ile Gln Val Ala Pro Ile Asp Ser
65 70 75

(2) INFORMATION FOR SEQ ID NO:29:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Leu Pro His Asn Phe Leu Thr Val Ala Pro Gly His Ser Ser His His
1 5 10 15
45 Ser Pro Gly Leu Gln Gly Gln Gly Val Thr Leu Pro Gly Glu Pro Pro
20 25 30
Leu Pro Glu Lys Lys Arg Val Ser Glu Gly Asp Arg Ser Leu Val Ser
35 40 45
50 Val Ser Pro Ser Ser Ser Gly Phe Ser Ser Pro His Ser Gly Ser Asn
50 55 60
55 Ile Ser Ile Pro Phe Pro Tyr Val Leu Pro Asp Phe Ser Lys Ala Ser
65 70 75 80
Glu Gly Gly Ser Thr Leu Gln Ile Val Gln Val Ile Asn Leu
85 90

(2) INFORMATION FOR SEQ ID NO:30:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 135 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Arg Arg Pro Pro Ala Asp Arg Gly Arg Ser Pro Pro Gly Gly Pro Gly
 1 5 10 15
 Ser Arg Thr Gly Glu Pro Gly Arg Glu Ser Ser Ala Ala Gly Cys Thr
 20 25 30
 Ala Ala Ala Pro Arg Glu Gly Cys Ser Gly Gln Arg Pro Pro Leu Leu
 25 35 40 45
 Arg Ala Asp Ser Ala Gly Leu Gly Arg Cys Gly Gly Leu Cys Arg Pro
 50 55 60
 Pro Val Ser Thr Tyr Cys Trp Arg Arg Phe Ala Pro Arg Pro Ala Glu
 65 70 75 80
 Trp Gly Gly Gly Pro Gly Arg Arg Thr Gly Gly Phe Ala Val Phe Leu
 85 90 95
 Gln Pro Pro Ile Leu Leu Ser Ser Pro Thr Ala Leu Gln Pro Ser Phe
 100 105 110
 Asp Asn Leu Arg Cys Leu Pro Ser Ser Ile His Cys Phe Gly Lys Gln
 40 115 120 125
 Pro Pro Ile Pro Pro Leu Leu
 130 135

45 (2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 937 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminal

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Leu Ala Ala Ala Gly Gly Arg Val Pro Thr Ala Ala Gly Ala Trp
 1 5 10 15
 5 Leu Leu Arg Gly Gln Arg Thr Cys Asp Ala Ser Pro Pro Trp Ala Leu
 20 25 30
 Trp Gly Arg Gly Pro Ala Ile Gly Gly Gln Trp Arg Gly Phe Trp Glu
 35 40 45
 10 Ala Ser Ser Arg Gly Gly Gly Ala Phe Ser Gly Gly Glu Asp Ala Ser
 50 55 60
 Glu Gly Gly Ala Glu Glu Gly Ala Gly Gly Ala Gly Gly Ser Ala Gly
 65 70 75 80
 15 Ala Gly Glu Gly Pro Val Ile Thr Ala Leu Thr Pro Met Thr Ile Pro
 85 90 95
 20 Asp Val Phe Pro His Leu Pro Leu Ile Ala Ile Thr Arg Asn Pro Val
 100 105 110
 Phe Pro Arg Phe Ile Lys Ile Ile Glu Val Lys Asn Lys Lys Leu Val
 115 120 125
 25 Glu Leu Leu Arg Arg Lys Val Arg Leu Ala Gln Pro Tyr Val Gly Val
 130 135 140
 Phe Leu Lys Arg Asp Asp Ser Asn Glu Ser Asp Val Val Glu Ser Leu
 145 150 155 160
 30 Asp Glu Ile Tyr His Thr Gly Thr Phe Ala Gln Ile His Glu Met Gln
 165 170 175
 35 Asp Leu Gly Asp Lys Leu Arg Met Ile Val Met Gly His Arg Arg Val
 180 185 190
 His Ile Ser Arg Gln Leu Glu Val Glu Pro Glu Glu Pro Glu Ala Glu
 195 200 205
 40 Asn Lys His Lys Pro Arg Arg Lys Ser Lys Arg Gly Lys Lys Glu Ala
 210 215 220
 Glu Asp Glu Leu Ser Ala Arg His Pro Ala Glu Leu Ala Met Glu Pro
 225 230 235 240
 Thr Pro Glu Leu Pro Ala Glu Val Leu Met Val Glu Val Glu Asn Val
 245 250 255
 50 Val His Glu Asp Phe Gln Val Thr Glu Glu Val Lys Ala Leu Thr Ala
 260 265 270
 Glu Ile Val Lys Thr Ile Arg Asp Ile Ile Ala Leu Asn Pro Leu Tyr
 275 280 285
 55 Arg Glu Ser Val Leu Gln Met Met Gln Ala Gly Gln Arg Val Val Asp
 290 295 300

	Asn Pro Ile Tyr Leu Ser Asp Met Gly Ala Ala Leu Thr Gly Ala Glu	
	305	310 315 320
5	Ser His Glu Leu Gln Asp Val Leu Glu Glu Thr Asn Ile Pro Lys Arg	
		325 330 335
	Leu Tyr Lys Ala Leu Ser Leu Leu Lys Lys Glu Phe Glu Leu Ser Lys	
		340 345 350
10	Leu Gln Gln Arg Leu Gly Arg Glu Val Glu Glu Lys Ile Lys Gln Thr	
		355 360 365
	His Arg Lys Tyr Leu Leu Gln Glu Gln Leu Lys Ile Ile Lys Lys Glu	
		370 375 380
15	Leu Gly Leu Glu Lys Asp Asp Lys Asp Ala Ile Glu Glu Lys Phe Arg	
		385 390 395 400
	Glu Arg Leu Lys Glu Leu Val Val Pro Lys His Val Met Asp Val Val	
		405 410 415
20	Asp Glu Glu Leu Ser Lys Leu Gly Leu Leu Asp Asn His Ser Ser Glu	
		420 425 430
25	Phe Asn Val Thr Arg Asn Tyr Leu Asp Trp Leu Thr Ser Ile Pro Trp	
		435 440 445
	Gly Lys Tyr Ser Asn Glu Asn Leu Asp Leu Ala Arg Ala Gln Ala Val	
		450 455 460
30	Leu Glu Glu Asp His Tyr Gly Met Glu Asp Val Lys Lys Arg Ile Leu	
		465 470 475 480
	Glu Phe Ile Ala Val Ser Gln Leu Arg Gly Ser Thr Gln Gly Lys Ile	
		485 490 495
35	Leu Cys Phe Tyr Gly Pro Pro Gly Val Gly Lys Thr Ser Ile Ala Arg	
		500 505 510
40	Ser Ile Ala Arg Ala Leu Asn Arg Glu Tyr Phe Arg Phe Ser Val Gly	
		515 520 525
	Gly Met Thr Asp Val Ala Glu Ile Lys Gly His Arg Arg Thr Tyr Val	
		530 535 540
45	Gly Ala Met Pro Gly Lys Ile Ile Gln Cys Leu Lys Lys Thr Lys Thr	
		545 550 555 560
	Glu Asn Pro Leu Ile Leu Ile Asp Glu Val Asp Lys Ile Gly Arg Gly	
		565 570 575
50	Tyr Gln Gly Asp Pro Ser Ser Ala Leu Leu Glu Leu Leu Asp Pro Glu	
		580 585 590
55	Gln Asn Ala Asn Phe Leu Asp His Tyr Leu Asp Val Pro Val Asp Leu	
		595 600 605
	Ser Lys Val Leu Phe Ile Cys Thr Ala Asn Val Thr Asp Thr Ile Pro	

	610	615	620
	Glu Pro Leu Arg Asp Arg Met Glu Met Ile Asn Val Ser Gly Tyr Val		
5	625	630	635 640
	Ala Gln Glu Lys Leu Ala Ile Ala Glu Arg Tyr Leu Val Pro Gln Ala		
	645	650	655
10	Arg Ala Leu Cys Gly Leu Asp Glu Ser Lys Ala Lys Leu Ser Ser Asp		
	660	665	670
	Val Leu Thr Leu Leu Ile Lys Gln Tyr Cys Arg Glu Ser Gly Val Arg		
	675	680	685
15	Asn Leu Gln Lys Gln Val Glu Lys Val Leu Arg Lys Ser Ala Tyr Lys		
	690	695	700
	Ile Val Ser Gly Glu Ala Glu Ser Val Glu Val Thr Pro Glu Asn Leu		
20	705	710	715 720
	Gln Asp Phe Val Gly Lys Pro Val Phe Thr Val Glu Arg Met Tyr Asp		
	725	730	735
25	Val Thr Pro Pro Gly Val Val Met Gly Leu Ala Trp Thr Ala Met Gly		
	740	745	750
	Gly Ser Thr Leu Phe Val Glu Thr Ser Leu Arg Arg Pro Gln Asp Lys		
	755	760	765
30	Asp Ala Lys Gly Asp Lys Asp Gly Ser Leu Glu Val Thr Gly Gln Leu		
	770	775	780
	Gly Glu Val Met Lys Glu Ser Ala Arg Ile Ala Tyr Thr Phe Ala Arg		
35	785	790	795 800
	Ala Phe Leu Met Gln His Ala Pro Ala Asn Asp Tyr Leu Val Thr Ser		
	805	810	815
40	His Ile His Leu His Val Pro Glu Gly Ala Thr Pro Lys Asp Gly Pro		
	820	825	830
	Ser Ala Gly Cys Thr Ile Val Thr Ala Leu Leu Ser Leu Ala Met Gly		
	835	840	845
45	Arg Pro Val Arg Gln Asn Leu Ala Met Thr Gly Glu Val Ser Leu Thr		
	850	855	860
	Gly Lys Ile Leu Pro Val Gly Gly Ile Lys Glu Lys Thr Ile Ala Ala		
50	865	870	875 880
	Lys Arg Ala Gly Val Thr Cys Ile Ile Leu Pro Ala Glu Asn Lys Lys		
	885	890	895
55	Asp Phe Tyr Asp Leu Ala Ala Phe Ile Thr Glu Gly Leu Glu Val His		
	900	905	910
	Phe Val Glu His Tyr Arg Glu Ile Phe Asp Ile Ala Phe Pro Asp Glu		
	915	920	925

Gln Ala Glu Ala Leu Ala Val Glu Arg
930 935

5 (2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 129 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

20 Thr Gly Glu Pro Cys Cys Asp Trp Val Gly Asp Glu Gly Ala Gly His
1 5 10 15

Phe Val Lys Met Val His Asn Gly Ile Glu Tyr Gly Asp Met Gln Leu
20 25 30

25

Ile Cys Glu Ala Tyr His Leu Met Lys Asp Val Leu Gly Met Ala Gln
35 40 45

30 Asp Glu Met Ala Gln Ala Phe Glu Asp Trp Asn Lys Thr Glu Leu Asp
50 55 60

Ser Phe Leu Ile Glu Ile Thr Ala Asn Ile Leu Lys Phe Gln Asp Thr
65 70 75 80

35 Asp Gly Lys His Leu Leu Pro Lys Ile Xaa Asp Ser Ala Gly Gln Lys
85 90 95

Gly Thr Gly Lys Trp Thr Ala Ile Phe Ala Leu Gly Leu Arg Gly Thr
100 105 110

40

Arg His Pro His Trp Gly Arg Cys Leu Xaa Ser Val Leu Ile Ile Ser
115 120 125

Xaa

45

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 376 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: peptide

55

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

5 Met Asp Met Val Glu Asn Ala Asp Ser Leu Gln Ala Gln Glu Arg Lys
 1 5 10 15
 Asp Ile Leu Met Lys Tyr Asp Lys Gly His Arg Ala Gly Leu Pro Glu
 20 25 30
 10 Asp Lys Gly Pro Glu Pro Val Gly Ile Asn Ser Ser Ile Asp Arg Phe
 35 40 45
 Gly Ile Leu His Glu Thr Glu Leu Pro Pro Val Thr Ala Arg Glu Ala
 50 55 60
 15 Lys Lys Ile Arg Arg Glu Met Thr Arg Thr Ser Lys Trp Met Glu Met
 65 70 75 80
 20 Leu Gly Glu Trp Glu Thr Tyr Lys His Ser Ser Lys Leu Ile Asp Arg
 85 90 95
 Val Tyr Lys Gly Ile Pro Met Asn Ile Arg Gly Pro Val Trp Ser Val
 100 105 110
 25 Leu Leu Asn Ile Gln Glu Ile Lys Leu Lys Asn Pro Gly Arg Tyr Gln
 115 120 125
 Ile Met Lys Glu Arg Gly Lys Arg Ser Ser Glu His Ile His His Ile
 130 135 140
 30 Asp Leu Asp Val Arg Thr Thr Leu Arg Asn His Val Phe Phe Arg Asp
 145 150 155 160
 35 Arg Tyr Gly Ala Lys Gln Arg Glu Leu Phe Tyr Ile Leu Leu Ala Tyr
 165 170 175
 Ser Glu Tyr Asn Pro Glu Val Gly Tyr Cys Arg Asp Leu Ser His Ile
 180 185 190
 40 Thr Ala Leu Phe Leu Leu Tyr Leu Pro Glu Glu Asp Ala Phe Trp Ala
 195 200 205
 Leu Val Gln Leu Leu Ala Ser Glu Arg His Ser Leu Pro Gly Phe His
 210 215 220
 45 Ser Pro Asn Gly Gly Thr Val Gln Gly Leu Gln Asp Gln Gln Glu His
 225 230 235 240
 50 Val Val Pro Lys Ser Gln Pro Lys Thr Met Trp His Gln Asp Lys Glu
 245 250 255
 Gly Leu Cys Gly Gln Cys Ala Ser Leu Gly Cys Leu Leu Arg Asn Leu
 260 265 270
 55 Ile Asp Gly Ile Ser Leu Gly Leu Thr Leu Arg Leu Trp Asp Val Tyr
 275 280 285
 Leu Val Glu Gly Glu Gln Val Leu Met Pro Ile Thr Ser Ile Ala Leu

290 295 300

Lys Val Gln Gln Lys Arg Leu Met Lys Thr Ser Arg Cys Gly Leu Trp
305 310 315 320

5 Ala Arg Leu Arg Asn Gln Phe Phe Asp Thr Trp Ala Met Asn Asp Asp
325 330 335

10 Thr Val Leu Lys His Leu Arg Ala Ser Thr Lys Lys Leu Thr Arg Lys
340 345 350

Gln Gly Asp Leu Pro Pro Pro Gly Pro Thr Ala Leu Gly Arg Arg Cys
355 360 365

15 Val Ala Gly Ser Pro Gln Pro Val
370 375

(2) INFORMATION FOR SEQ ID NO:34:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

35 Glu Phe Cys Gly Arg Gln Lys Ile His Lys Glu Met Pro Cys Lys Cys
1 5 10 15

Thr Val Cys Gly Ser Asp Phe Cys His Thr Ser Tyr Leu Leu Glu His
20 25 30

40 Gln Arg Val His His Glu Glu Lys Ala Tyr Glu Tyr Asp Glu Tyr Gly
35 40 45

Leu Ala Tyr Ile Lys Gln Gln Gly Ile His Phe Arg Glu Lys Pro Tyr
50 55 60

45 Thr Cys Ser Glu Cys Gly Lys Asp Phe Arg Leu Asn Ser His Leu Ile
65 70 75 80

Gln His Gln Arg Ile His Thr Gly Glu Lys Ala His Glu Cys His Glu
85 90 95

50 Cys Gly Lys Ala Phe Ser Gln Thr Ser Cys Leu Ile Gln His His Lys
100 105 110

55 Met His Arg Lys Glu Thr Arg Ile Glu Cys Asn Glu Tyr Xaa Gly Gln
115 120 125

Val Gln Val Ile Ala Gln Ile Leu Ser Cys Asn Lys Glu Val Leu Thr
130 135 140

Arg Gln Lys Ala Phe Asp Trp Xaa Cys Met Gly Lys Glu Leu Gln Ser
 145 150 155 160
 5 Glu Ser Thr Ser Ser Ser Thr Ser Glu His Ser Tyr Gln Arg Glu Leu
 165 170 175
 Met Asn Val Met Lys Met Gly Arg Tyr Leu Ser Asn Ser Gly Phe Ile
 180 185 190
 10 Gln His Leu Arg Val His Thr Arg Glu Gln Ile Met Tyr Val Leu His
 195 200 205
 Val Val Lys Pro Ser Val Ile Ala Gln Pro Leu Leu Ser Ile Arg Xaa
 210 215 220
 15 Phe Thr Pro Glu Arg Asn Pro Leu Asn Val Thr Asn Glu Glu Lys Val
 225 230 235 240
 20 Leu Val Leu Asn Ser Xaa Ser Thr Pro Ala Asn Leu Tyr Gln Xaa Glu
 245 250 255
 Ile Leu Gln Met Tyr Trp Ile Val Ala Asn Phe Ser Cys Tyr Xaa Tyr
 260 265 270
 25 Phe His Thr Leu Val Thr Cys Gly Gly Ile His Met Gly Ile Asn Ser
 275 280 285
 His Cys Cys Asn Asp Cys Glu Lys His Gln Ala Arg Asn Phe Leu Val
 290 295 300
 30 Arg Phe Asn Ser Thr Pro Cys Lys Arg Phe Leu
 305 310 315
 35 (2) INFORMATION FOR SEQ ID NO:35:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 127 amino acids
 (B) TYPE: amino acid
 40 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: internal
 45
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:
 50 Leu Phe Ala Glu Ala Gly Pro Asp Phe Glu Leu Arg Leu Glu Leu Tyr
 1 5 10 15
 Gly Ala Cys Val Glu Glu Glu Gly Ala Leu Thr Gly Gly Pro Lys Arg
 20 25 30
 55 Leu Ala Thr Lys Leu Ser Ser Ser Leu Gly Arg Ser Ser Gly Arg Arg
 35 40 45

Val Arg Ala Ser Leu Asp Ser Ala Gly Gly Ser Gly Ser Ser Pro Ile
 50 55 60

Leu Leu Pro Thr Pro Val Val Gly Gly Pro Arg Tyr His Leu Leu Ala
 5 65 70 75 80

His Thr Thr Leu Thr Leu Gly Gly Val Gln Asp Gly Phe Arg Thr His
 85 90 95

Asp Leu Thr Leu Gly Ser His Glu Glu Asn Leu Pro Gly Cys Pro Phe
 10 100 105 110

Met Val Ala Cys Val Ala Val Trp Gln Leu Ser Leu Ser Ala Xaa
 15 115 120 125

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 278 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

His Glu Ser Lys Gln Glu Lys Glu Lys Ser Lys Lys Lys Lys Gly Gly
 1 5 10 15

Lys Thr Glu Gln Asp Gly Tyr Gln Lys Pro Thr Asn Lys His Phe Thr
 35 20 25 30

Gln Ser Pro Lys Glu Val Ser Gly Arg Pro Ala Gly Val Leu Trp Lys
 35 40 45

Ala Asn Glu Gly Leu Leu Leu Ile Thr Ala Pro Lys Ala Glu Glu Gln
 40 50 55 60

Gln Arg Asp Glu Tyr Leu Glu Ser Phe Cys Lys Met Ala Thr Arg Lys
 45 65 70 75 80

Ile Ser Val Ile Thr Ile Phe Gly Pro Val Asn Asn Ser Thr Met Lys
 85 90 95

Ile Asp His Phe Gln Leu Asp Asn Glu Lys Pro Met Arg Val Val Asp
 50 100 105 110

Asp Glu Asp Leu Val Asp Gln Arg Leu Ile Ser Glu Leu Arg Lys Glu
 115 120 125

Tyr Gly Met Thr Tyr Asn Asp Phe Phe Met Val Leu Thr Asp Val Asp
 55 130 135 140

Leu Arg Val Lys Gln Tyr Tyr Glu Val Pro Ile Thr Met Lys Ser Val

[illegible]

(2) INFORMATION FOR SEQ ID NO:37:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 292 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

45 His Tyr Ser Cys Asn Ile Ser Gly Ser Leu Lys Arg His Tyr Asn Arg
1 5 10 15
Lys His Pro Asn Glu Glu Tyr Ala Asn Val Gly Thr Gly Glu Leu Ala
20 25 30
Ala Glu Val Leu Ile Gln Gln Gly Gly Leu Lys Cys Pro Val Cys Ser
35 40 45
Phe Val Tyr Gly Thr Lys Trp Glu Phe Asn Arg His Leu Lys Asn Lys
50 55 60
55 His Gly Leu Lys Val Val Glu Ile Asp Gly Asp Pro Lys Trp Glu Thr
65 70 75 80
Ala Thr Glu Ala Pro Glu Glu Pro Ser Thr Gln Tyr Leu His Ile Thr

	85	90	95
	Glu Ser Glu Glu Asp Val Gln Gly Thr Gln Ala Ala Val Ala Ala Leu		
	100	105	110
5	Gln Asp Leu Arg Tyr Thr Ser Glu Ser Gly Asp Arg Leu Asp Pro Thr		
	115	120	125
10	Ala Val Asn Ile Leu Gln Gln Ile Ile Glu Leu Gly Ala Glu Thr His		
	130	135	140
	Asp Ala Thr Ala Leu Ala Ser Val Val Ala Met Ala Pro Gly Thr Val		
	145	150	155
15	Thr Val Val Lys Gln Val Thr Glu Glu Glu Pro Ser Ser Asn His Thr		
	165	170	175
	Val Met Ile Gln Glu Thr Val Gln Gln Ala Ser Val Glu Leu Ala Glu		
	180	185	190
20	Gln His His Leu Val Val Ser Ser Asp Asp Val Glu Gly Ile Glu Thr		
	195	200	205
25	Val Thr Val Tyr Thr Gln Gly Gly Glu Ala Ser Glu Phe Ile Val Tyr		
	210	215	220
	Val Gln Glu Ala Met Gln Pro Val Glu Glu Gln Ala Cys Gly Ala Ala		
	225	230	235
30	Gly Pro Gly Thr Leu Glu Asp Met Trp His Arg Met Ala Thr Gly Arg		
	245	250	255
	Gly Cys Pro Gly Ser Ser Gly Thr Gln Gly Gly Glu Ala Thr Phe Leu		
	260	265	270
35	Pro Tyr Pro Arg Met Val Ser Pro Leu Pro Ser Leu Pro Ser Ser Leu		
	275	280	285
40	Ile Gly Leu Ser		
	290		

(2) INFORMATION FOR SEQ ID NO:38:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 83 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: internal

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg

1 5 10 15
 Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
 20 25 30
 5 Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
 35 40 45
 10 Glu Arg Glu Arg Glu Ser Pro Gly Leu Asn Thr Tyr Gly Thr Asp Val
 50 55 60
 Ile Ser Thr Ser Pro Phe Ile Glu Ser Val Ile Tyr Leu Glu Trp Arg
 65 70 75 80
 15 His Arg Phe

(2) INFORMATION FOR SEQ ID NO:39:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 191 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 25 (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: internal
 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
 Glu Phe Cys Gly Arg Arg Ser Glu Val Leu Leu Val Ser Glu Asp Gly
 1 5 10 15
 35 Lys Ile Leu Ala Glu Ala Asp Gly Leu Ser Thr Asn His Trp Leu Ile
 20 25 30
 40 Gly Thr Asp Lys Cys Val Glu Arg Ile Asn Glu Met Val Asn Arg Ala
 35 40 45
 Lys Arg Lys Ala Gly Val Asp Pro Leu Val Pro Leu Arg Ser Leu Gly
 50 55 60
 45 Leu Ser Leu Ser Gly Gly Asp Gln Glu Asp Ala Gly Arg Ile Leu Ile
 65 70 75 80
 Glu Glu Leu Arg Asp Arg Phe Pro Tyr Leu Ser Glu Ser Tyr Leu Ile
 85 90 95
 50 Thr Thr Asp Ala Ala Gly Ser Ile Asp Thr Ala Thr Pro Asp Gly Gly
 100 105 110
 55 Val Val Leu Ile Ser Gly Thr Gly Ser Asn Cys Arg Leu Ile Asn Pro
 115 120 125
 Asp Gly Ser Glu Ser Gly Cys Gly Arg Leu Gly Gly Ile Leu Trp Val
 130 135 140

Met Arg Val Gln Pro Thr Gly Ser His Thr Lys Gln Xaa Lys Xaa Cys
 145 150 155 160

5 Leu Asp Ser Ile Glu Asn Xaa Arg Arg Ser His Asp Ile Gly Tyr Val
 165 170 175

Lys Gln Ala Met Phe His Tyr Phe Gln Val Gln Ile Arg Xaa Val
 180 185 190

10

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 56 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: peptide

20

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

25

Gln Ser Ser Thr Glu Ile Ser Lys Thr Arg Gly Gly Glu Thr Lys Arg
 1 5 10 15

30

Glu Val Arg Val Glu Glu Ser Thr Gln Val Gly Gly Ala Pro Leu Pro
 20 25 30

Cys Cys Val Trp Gly Leu Pro Gly Pro Gly Ala Pro Gly Ile Leu Arg
 35 40 45

35

Gln Tyr His Pro Ala Ala Gly Gly
 50 55

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 93 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: peptide

45

(v) FRAGMENT TYPE: internal

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

55

Gly Glu Glu Lys Arg Val Ser Arg Glu Pro Ala Gly Val Leu Ser Gln
 1 5 10 15

Ser Gly Met Gln Leu Glu Tyr Leu Ser Leu Pro Phe Gln Leu Pro Ala
 20 25 30

Arg Arg Ser Leu Gln Val Glu Leu Cys Gly Gly Gln Pro Val Leu Ser
 35 40 45
 5 Arg Val Lys Val Gln Trp Arg Pro Ser Gly Ser Thr Pro Asn Val Ile
 50 55 60
 Glu Gly Asp Leu Leu Val Phe Gly Gln Gln Leu Ala Pro Pro Met Gly
 65 70 75 80
 10 Met Gly Glu Val Met Glu Glu Glu Arg Arg Leu Cys Xaa
 85 90

(2) INFORMATION FOR SEQ ID NO:42:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 84 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 20 (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: internal
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
 Ala Leu His Val Asn Asn Asp Arg Ala Lys Val Ile Leu Lys Pro Asp
 1 5 10 15
 30 Lys Thr Thr Ile Thr Glu Pro His His Ile Trp Pro Thr Leu Thr Asp
 20 25 30
 Glu Glu Trp Ile Lys Val Glu Val Gln Leu Lys Asp Leu Ile Leu Ala
 35 35 40 45
 Asp Tyr Gly Lys Lys Asn Asn Val Asn Val Ala Ser Leu Thr Gln Ser
 50 55 60
 40 Glu Ile Arg Asp Ile Ile Leu Gly Ile Glu Asp Leu Arg Glu Pro Ser
 65 70 75 80
 Gln Glu Gly Glu

45

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 382 amino acids
 50 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: peptide
 55 (v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

5	Met Ser Asp Ser Lys Cys Asp Ser Gln Phe Tyr Ser Val Gln Val Ala	1 5 10 15
	Asp Ser Thr Phe Thr Val Leu Lys Arg Tyr Gln Gln Leu Lys Pro Ile	20 25 30
10	Gly Ser Gly Ala Gln Gly Ile Val Cys Ala Ala Phe Asp Thr Val Leu	35 40 45
	Gly Ile Asn Val Ala Val Lys Lys Leu Ser Arg Pro Phe Gln Asn Gln	50 55 60
15	Thr His Ala Lys Arg Ala Tyr Arg Glu Leu Val Leu Leu Lys Cys Val	65 70 75 80
	Asn His Lys Asn Ile Ile Ser Leu Leu Asn Val Phe Thr Pro Gln Lys	85 90 95
20	Thr Leu Glu Glu Phe Gln Asp Val Tyr Leu Val Met Glu Leu Met Asp	100 105 110
	Ala Asn Leu Cys Gln Val Ile His Met Glu Leu Asp His Glu Arg Met	115 120 125
25	Ser Tyr Leu Leu Tyr Gln Met Leu Cys Gly Ile Lys His Leu His Ser	130 135 140
	Ala Gly Ile Ile His Arg Asp Leu Lys Pro Ser Asn Ile Val Val Lys	145 150 155 160
30	Ser Asp Cys Thr Leu Lys Ile Leu Asp Phe Gly Leu Ala Arg Thr Ala	165 170 175
35	Cys Thr Asn Phe Met Met Thr Pro Tyr Val Val Thr Arg Tyr Tyr Arg	180 185 190
	Ala Pro Glu Val Ile Leu Gly Met Gly Tyr Lys Glu Asn Val Asp Ile	195 200 205
40	Trp Ser Val Gly Cys Ile Met Gly Glu Leu Val Lys Gly Cys Val Ile	210 215 220
	Phe Gln Gly Thr Asp His Ile Asp Gln Trp Asn Lys Val Ile Glu Gln	225 230 235 240
45	Leu Gly Thr Pro Ser Ala Glu Phe Met Lys Lys Leu Gln Pro Thr Val	245 250 255
50	Arg Asn Tyr Val Glu Asn Arg Pro Lys Phe Pro Gly Ile Lys Leu Glu	260 265 270
	Glu Leu Phe Pro Asp Trp Leu Phe Pro Ser Glu Ser Glu Arg Asp Lys	275 280 285
55	Ile Lys Thr Ser Gln Ala Arg Asp Leu Leu Ser Gln Met Leu Val Ile	290 295 300

Asp Pro Asp Lys Arg Ile Ser Val Asp Glu Ala Leu Arg His Pro Tyr
 305 310 315 320
 5 Ile Thr Val Trp Tyr Asp Pro Ala Glu Ala Glu Ala Pro Pro Pro Pro
 325 330 335
 Ile Tyr Asp Ala Gln Leu Glu Glu Arg Glu His Ala Ile Glu Glu Trp
 340 345 350
 10 Lys Glu Leu Ile Tyr Lys Glu Val Met Asp Trp Glu Glu Arg Ser Lys
 355 360 365
 Asn Gly Val Val Lys Asp Gln Pro Ser Ala Gln Met Gln Gln
 15 370 375 380

(2) INFORMATION FOR SEQ ID NO:44:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 151 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: internal

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

His Glu Glu Asn Met His Asp Leu Gln Tyr His Thr His Tyr Ala Gln
 1 5 10 15
 35 Asn Arg Thr Val Glu Arg Phe Glu Ser Leu Val Gly Arg Met Ala Ser
 20 25 30
 His Glu Ile Glu Ile Gly Thr Ile Phe Thr Asn Ile Asn Ala Thr Asp
 40 35 40 45
 Asn His Ala His Ser Met Leu Met Tyr Leu Asp Asp Val Arg Leu Ser
 50 55 60
 45 Cys Thr Leu Gly Phe His Thr His Ala Glu Glu Leu Tyr Tyr Leu Asn
 65 70 75 80
 Lys Ser Val Ser Ile Met Leu Gly Thr Thr Asp Leu Leu Arg Glu Arg
 85 90 95
 50 Phe Ser Leu Leu Ser Ala Arg Leu Asp Leu Asn Val Arg Asn Leu Ser
 100 105 110
 Met Ile Val Glu Glu Met Lys Gly Gly Asp Thr Gln Asn Gly Glu Ile
 55 115 120 125
 Leu Arg Asn Val Thr Ser Tyr Glu Val Pro Pro Ala Ser Arg Thr Lys
 130 135 140

Arg Phe Lys Arg Asp Leu Ala
145 150

5 (2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 373 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

15 (v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

20 Met Val Asp Tyr Ser Val Trp Asp His Ile Glu Val Ser Asp Asp Glu
1 5 10 15

Asp Glu Thr His Pro Asn Ile Asp Thr Ala Ser Leu Phe Arg Trp Arg
20 25 30

25 His Gln Ala Arg Val Glu Arg Met Glu Gln Phe Gln Lys Glu Lys Glu
35 40 45

30 Glu Leu Asp Arg Gly Cys Arg Glu Cys Lys Arg Lys Val Ala Glu Cys
50 55 60

Gln Arg Lys Leu Lys Glu Leu Glu Val Ala Glu Gly Gly Lys Ala Glu
65 70 75 80

35 Leu Glu Arg Leu Gln Ala Glu Ser Thr Ala Ala Ala Gln Gly Gly Ala
85 90 95

Glu Leu Gly Ala Glu Ala Gly Gly Arg Cys Ala Arg Arg Arg Ala
100 105 110

40 Cys Pro Gly Asn Val Asp Thr Leu Ser Lys Asp Gly Phe Ser Lys Ser
115 120 125

Met Val Asn Thr Lys Pro Glu Lys Thr Glu Glu Asp Ser Glu Glu Val
45 130 135 140

Arg Glu Gln Lys His Lys Thr Phe Val Glu Lys Tyr Glu Lys Gln Ile
145 150 155 160

50 Lys His Phe Gly Met Leu Arg Arg Trp Asp Asp Ser His Lys Tyr Leu
165 170 175

Ser Asp Asn Val His Leu Val Cys Glu Glu Thr Ala Asn Tyr Leu Val
180 185 190

55 Ile Trp Cys Ile Asp Leu Glu Val Glu Glu Lys Cys Ala Leu Met Glu
195 200 205

Gln Val Ala His Gln Thr Ile Val Met Gln Phe Ile Leu Glu Leu Ala
 210 215 220
 5 Lys Ser Leu Lys Val Asp Pro Arg Ala Cys Phe Arg Gln Phe Phe Thr
 225 230 235 240
 Lys Ile Lys Thr Ala Asp Arg Gln Tyr Met Glu Gly Phe Asn Asp Glu
 245 250 255
 10 Leu Glu Ala Phe Lys Glu Arg Val Arg Gly Arg Ala Lys Leu Arg Ile
 260 265 270
 Glu Lys Ala Met Lys Glu Tyr Glu Glu Glu Glu Arg Lys Lys Arg Leu
 275 280 285
 15 Gly Pro Gly Gly Leu Asp Pro Val Glu Val Tyr Glu Ser Leu Pro Glu
 290 295 300
 20 Glu Leu Gln Lys Cys Phe Asp Val Lys Asp Val Gln Met Leu Gln Asp
 305 310 315 320
 Ala Ile Ser Lys Met Asp Pro Thr Asp Ala Lys Tyr His Met Gln Arg
 325 330 335
 25 Cys Ile Asp Ser Gly Leu Trp Val Pro Asn Ser Lys Ala Ser Glu Ala
 340 345 350
 Lys Glu Gly Glu Glu Ala Gly Pro Gly Asp Pro Leu Leu Glu Ala Val
 355 360 365
 30 Pro Lys Thr Gly Arg
 370

35 (2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 143 amino acids
 (B) TYPE: amino acid
 40 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

50 Arg Arg His Pro Ser Arg Ser Gly Leu Gly Arg Gln Gly Lys Met Val
 1 5 10 15
 Asp Tyr Ser Val Trp Asp His Ile Glu Val Ser Asp Asp Glu Asp Glu
 20 25 30
 55 Thr His Pro Asn Ile Asp Thr Ala Ser Leu Phe Arg Trp Arg His Gln
 35 40 45

Ala Arg Val Glu Arg Met Glu Gln Phe Gln Lys Glu Lys Glu Glu Leu
 50 55 60

5 Asp Ser Gly Cys Arg Glu Cys Lys Arg Lys Val Ala Glu Cys Gln Arg
 65 70 75 80

Lys Leu Lys Glu Leu Glu Val Ala Glu Gly Gly Lys Ala Glu Leu Glu
 85 90 95

10 Arg Leu Gln Ala Glu Ala Gln Gln Leu Arg Asn Glu Glu Arg Ser Trp
 100 105 110

Glu Gln Lys Leu Glu Glu Met Arg Lys Lys Glu Lys Ser Met Pro Trp
 115 120 125

15 Gln Arg Gly His Ala Gln Gln Arg Arg Leu Gln Gln Arg Ala Trp
 130 135 140

(2) INFORMATION FOR SEQ ID NO:47:

20

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 77 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

25

- (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: internal

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

35 Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
 1 5 10 15

Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg
 20 25 30

40 Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Arg Glu Ser Leu Tyr
 35 40 45

Asp Leu Ser Ile Gln Asn Phe Gln Val Ser Pro Tyr Val Glu Pro Lys
 50 55 60

45 Ser Phe Phe Leu Pro Arg Asn Phe Thr Thr Ile Arg Xaa
 65 70 75

(2) INFORMATION FOR SEQ ID NO:48:

50

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 72 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

55

- (ii) MOLECULE TYPE: peptide
 (v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

5 Met Ser Asp Ser Asn Cys Glu Ser Gln Phe Phe Gly Val Lys Val Glu
1 5 10 15

10 Asp Ser Thr Ser Thr Val Leu Lys Arg Tyr Gln Lys Leu Lys Pro Ile
20 25 30

Gly Ser Gly Ala Gln Gly Ile Val Gly Ala Ala Ser Gly Thr Val Leu
35 40 45

15 Gly Asp Lys Cys Trp Ser Gln Gly Ile Lys Pro Ala Pro Phe Gln Asn
50 55 60

20 Pro Thr His Glu Arg Glu Phe Ser
65 70

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 548 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

35 CCCAGGTTTA ATGATTTTATT TAACTGGTGG GAACAAAAAT TAACCCAGAT TACCCACACC 60

CATGCCTAAC TTTATCAATT GTTTAGGAGG TAATTTTGAT TCTTATTGA AAAAATGTTC 120

40 CATCCATTAT AAACAATTCC CAATAATCCG GTCAATTATT TTCCTAAAT TCCCCCAAT 180

TCCTTAGGAG AGGATGTAAT TGGGAGGTAA CTTTGGACG GCTTACTATC TTAACAAGNT 240

45 TGGGGTGAAG GGTGAGGAG TCCAAACCCT TCCAGATGG TGGGNGNNGG GTNAAGGAAT 300

TCCCTTTNTC CCCCCCCCCC NNNGGGGNCN GCCCCCCCCC NGGNNCCCC CNGGGGGGAA 360

CCCNCTCCNG TTTNAANAAA AAANNNGGGG GAGAGNCNA NAGCGGGGGT TTTTITGGG 420

50 GGGCCCCCCC CCCCCNCCN AAANTTCTCC CCCCCNAGNG GGGGAAANNG NCNNCNCNTT 480

TTCACINCA CNNTNCNCC NGCINNNGGGG GGGGGTTCC CCCCCCNC NCGGNNCCCC 540

55 CCCCCCCC 548

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 239 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TCCCCCAAGT CCAAATTTTT TTTTCTCT GATTGGGGAT GATTTTTAGG GGGAAGGGAA 60
ATTGATTTTC AAAAGGTTTT TTGAAAATC CATTTAAATC CTGGTTTTTT CCTTAAAAGT 120
TTCAGAAAGG TAAAATTTTG AACTAAAAAG GAAGGGAGGC CGTAACAAGG TTTTGGGTGT 180
TGAGATTAAT TGAACAGGGA TTTTAAACAT GGTTTTGGTT TACAACTGGG GGAATANAA 239

20

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 379 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: cDNA

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GGGTGATCAT GCACAAGTCT TAATTTATTG GGTAAAAACA TTAATTTATT ACAACATTTT 60
TCCAATAAA GCATAATAAA TAGAATCCAT TTCTTTTAAA ACGCTGTACA AGAGACTGGA 120
AAACAAGCTC CCAACAGAAT ATGAATAACT CATAACTCAT CCTACCTTCT TATTGATTGG 180
GGACGCTCCC CCCACCCCCC ATGCCTGAAG CAACGTGCAC ACTTCAGGTC TCTGARCACA 240
GCCGGCCAAG GCCACCAGCT TCTAGGSTCC CTGGAGGTCA TGACTTCACT CTTAAATGCT 300
CTGCCCTTGG GTCTCGTCTT AGGCCAGGA GGCTGAGGGC AGGAGAAGTG ACCCGTTAGG 360
TGTTGTGGC CTGGAGGAG 379

45

(2) INFORMATION FOR SEQ ID NO:52:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 296 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

5 ATCAGTCTGA TGTAGCTTTT ATTGAGTAAA GGAAAAAGGG AATTCAGCCG CATGATACAG 60
AGGTTCCAGT TGATCAGAGT GCGCAAACAC CCTTCCTGTC TGCCTGATGG GAACCGCACC 120
AGCACACGGG GTACGCGGAA GCCACTGCCG CAAGGAGATG GTTCCCACTC TCACGCACAT 180
10 GAGCAGCTCC TGGTCAGTCC CAAGAGGCAA GGGCAGAGGG CATGGTGGCT CTCACAGAGC 240
TACTTTACAA ATAACTGTG TGTCTTCCTC AGGAGTCTCT TACAACACTT TAAAAA 296

(2) INFORMATION FOR SEQ ID NO:53:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 365 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

AACTATTTTA ATTAGAATTT TTATTTGGTG CTTCAGGGCC ACAGGATAAA ATAACTACAT 60
30 TTAGCTTGCC TTTCAGTGAC GCTTTGGCCA AATGTCAGCT ACAAGGAGTC ATCTCCCTCA 120
CCGCCAAGCT GTCTAGCAGC CAGAGTGGTA GCTTTACTGT AACACACAGT ACTTTTGGTA 180
ATCAGACTCA AAGTCTTCAT CCATACTGCT TGTGTCTGCC ATCTTTTGGG CATCAGTCTT 240
35 GGGCAGAAAT TGTGCATAGT CTATCCCCTG CTGCTCATAG AAAAGATTGT AGGCAGAGTC 300
GGGTGTCAAT TTCATCCGGG TGAAGTTCCT TACAGCTGCT GTCATTGTAC AAGTACCACT 360
40 TGCAG 365

(2) INFORMATION FOR SEQ ID NO:54:

45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 339 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

55 CCAGAATACC AAACACACCT TTATCCAGGT GGAAGTACAA AAGCACATCC CTAAACCAAA 60
CGCATACATG TGATTTTAC ATTTCCTGTT TTTTAGGGAT TACATAATCC TGTTTCAGTC 120

ACCATACGTG ACTACTGGTC TCTATACATA AGGGTATACA TGTTGGACAG GAAAAAACAC 180
 ATGCATTTTC CATTGGCTTT TACATTTRGA TCACTCCATT TATTTTTCOA TTTCATTTAG 240
 ATTCCTACCT GGCCTGGATG AAATCCTACT CTGCTGATG GCAAAGAAGT AAAATATAGT 300
 GGCAGAACTA TCCTAGAGGG TTAGCCATAG GGGGATTAT 339

10 (2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 529 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

AGCCATAGGA GTTATAGAGT GAGCAACATA TTTGTATGTA TTTGTTGAGG GTCCCTACTG 60
 AATATTATAA CACTGCAACT ATGAAAGCCT CAATTGCTGG ACTGACAACA AGAATTTTAA 120
 ATAACATTTG TCTTACTCAC AAAATGTTAT AAAGCTTAAG ATGGAAAAAT ACAAATGTT 180
 GGGACATTAC CTAAGAATC ATGAACTCTT GTTAGGTATA TGATGGTGGC CCTGAAGTTG 240
 AGCCAACATC TTGTAATCAC TTTTATCAGT CAAAAAGCCA TGTTCTTTTA TATAGCCTGT 300
 AGACTATTAA AATACAAAAA TGTGGTAATG GATAACAAC TATACACAAA GCCCTCACAC 360
 TTCAAATACT GTCCTGGATT GATGAGAGAG GAGCAGAATT CAACCATTTA TCTGCAATCC 420
 TAATGGGTAA AATTTTACCA GGAACAGACC TGCCTCTCT GAATACTGCT CTGAGATTAC 480
 ATACGACAGG ATCATCTCTT GTTGGGAGGC TACATCCCCT ATGAGCGAT 529

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 386 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGCTGTAA TAACTTTAAT GGTGATGTG GGAGTCACAA GGGAGGTATG TTGGCTCCAA 60
 GGGTCTCCA GTGCCATCCT CAAAGCTGGT TAGTGAAGGG AGGTAGGGAA GAGTTGGTTC 120

CAGTTTTCTC CCAGGAAGGG TTAGGGAGG TCCCAGCGAG CCCCAGGAAT GAGTCCCTCG 180
 GTACCATGGA AACCACAATT TAAGAGGGGC TTCTGCCCAC CCCTGCAGCC TACCCCAGGT 240
 5 CCAGCAGAGG AACAGGAGGC CAGACTGGCC AACTTGCTAT AGACAGCGCC GTATCCAGAG 300
 CCCAACTGCG CATGGGTCAT TTTCTCTTCT GGGCAGATCC TATGCCAGAC CTTCTCTCTC 360
 10 ACACTGGTGA CTTGGAGCCA AGTGCG 386

(2) INFORMATION FOR SEQ ID NO:57:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 306 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

25 AAGGTGAAAG TTGGCTATTT ATTTAGTCTT AGAAAAACAC TGAAAGAAAA AGGCAGGAAA 60
 TGTAGTACGC AGTGTGGGAA GAATGGGGGC TGGCCACATG TAGTTTTAGC AAGCTGCAGA 120
 30 GGAAACCTGG CTGAGTTCTA AGGTTACAAT TTTTCTTGTT CAGGAAGGGG TTTCCAAGGG 180
 GAATACCTCT CATGATGGAC GGGAGCCAAT CCCGGTAACC CACCCCGGGT TTCCCGGGGG 240
 GGTAACCTTG GGAAACCCAT GGCCTGGAAT CCTCATCTTT CCTGGGAAGG GGCATCCCCA 300
 35 GGGGAA 306

(2) INFORMATION FOR SEQ ID NO:58:

40 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 471 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

50 CTGCGAAAGC CGAACTTTTT TGGGGGTTTC CCACCTAAGA AGTTCCCAGT TGAGTTGAAT 60
 GAAATGTGAA AAAGTCCCCT AGAAAGTTGG GCCTCGCAGT GTGTAAAAAA GGCCCCCAT 120
 55 GGGGAAGAGC CGTGAAACCA TTTTAAAAAA AGAGAAAGTG AGAGAGAATT CAGGCCCCCT 180
 GGGAGCCTGG TTTGGGTGGA GTGAACATCG TTCAGGCCGG CCCATGTGCC AGGCCACTCC 240

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

TGTTGGTTCG GGGGCTGTTT TCTTCTCTAA TTGTGCTTTC CCNNCCAAGT CCTAAAAANCT 300
CTGGGGTTGN GGCCACCAGA NAGACCAGAC CAANTCCCCG GGTNAAGAG GGTTTNTTNC 360
CTNGGCGAAG TTGGNGGTGC CCCAAAAAAG NNACCCNAAA AANTNTTCCC CCCTTTCAGC 420
CCCCNGANN CAAGGTTCCT TGGCNGANC CCCCAACCT NTTCCCACC C 471

10 (2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 463 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

ATACAAAATT TATTATTATA TTTTATTCAG GATGACAAGC CATCAGGAGG TCAACAACAC 60
AAGCACAGAC AGAGGGAAAG AGGGCAACCT GCTGAATGTC AGGGGCTGTC TTGAGGGGTT 120
GAGGGTTCCG CCCTCGGGAG GGTGAGGAA GAGGGAAGGG AACCGGCAAG GATTCAAGTT 180
CCCCCCTCC CGAGGGGTAA CCCTCCCCTC CTAAGGAGAA AAGTTGAGGG ATGTGAGAGG 240
CCTTTAACCC GTGCGGAGAT CTCTGTGGTG CCCCCCAGT TGGNCTCATT TNCATTGGG 300
GGACAACCCC CACACCCATA NGNTNGNNGT NCCNCGNGG TCTTNGAGG NCCNTNNGG 360
NCGCCAAGGA ANNGCCCCAA AAGAAGATNT TCACCCTNTC ATTGNTTNA GGAAGTCCCN 420
TGGGNNNGC CGCCTCTTTT TTCNTTGGG CCCCTCCCN CCC 463

40 (2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 392 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

GAATTCGGCA CGAGGTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTGAAT GGGGTTATCC 60
AGGATGTGAC TTTGGGAGAT TGGTTTTTTC CGTGGATTAT CCTGCCCCCTG AGATCCACCC 120
AAGTTGTGGG ATCTGAAACT GGCCACCCCT CCGGGATTTT GAAGGACGCT GAATCATGAG 180

CGACAGTAAT TGTGAAAGCC AGTTTTTTGG TGTGAAAGTG GAAGACTCAA CCTCCACTTG 240
TCCTAAAACG GTTACCAGAA GTTGAACCCA ATTGGTTCCT GGGGCCCAAG GGATTGTTGG 300
GTGTTGCATT GGGTACAGCC CTTGGGATAA TTGTTGGAGG CCAAGAAATT AGGCCCCCCT 360
TTCCAGACCC AACTCATGAA AGGGAGTTCT CC 392

10 (2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 506 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

TTGACCAAAC CTCTGGCGAA GAAGTCCAAA GCTTCTCGAG GGCCAACAGG GCCCCTTTCT 60
CCCACAGGCC CGGCCTCTCC AGGTTGTCCC TGAGGACCCT GGGGTCCCAG GGGGCCCAAG 120
CTGCCGGGGT CTCCTTTCGG GCCTCTGCCG CCAACAGGCC CTTTCACGCC CATATCTCCT 180
TGGAATCCTC TTGGTCCTGG AGGGCCGGGG GCACCTCGTA GGATGGTGAC ATTGCGAAGG 240
ATTTCTCCAT GCTGTGTGTC CACTGCCTTC ATCTCCTCCA CGATCATGGA GAGGTTCCGG 300
ACGTTGAGGT CCAGCCGGGC ACTGAGCAGG CTGAAGCGCT CCCGGAGCAG GTCTGTGGTG 360
CCCAGCATGA TGGAGACAGA CTTGTTTCAGG TAGTAGAGCT CCTCGGCATG GGTGTGGAAG 420
CCCAGCGTGC AGGAGAGCCG AACGTCATCC AGGTACTTGG AGCATGTTGT GCACGTGGTG 480
GTCGGTGGAA TTGATGTTGG TGAAGA 506

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCAAAGGCAT TCAGGCTCTT TAATGTCTGA GGATGGGGGG AAGAAGTCAA TGGTGAGGCT 60
CCTCTGGGAA ATTCTGAAGG CCTGGTGGTT CTCTAAGCCC CTCTAGCAAC ATGTGGATAT 120

GGGCTTGGAT ATCCATGGAG TCCTTGGTGA GGCTGTTGCT GAGCTCTGTG AGGAGAGAGC 180
TCTTACGACC AATGAACTGG AGAGCTTCTG CCAGTGTCAC CTCCAGGAAA AAACCATATC 240
5 CCAGGGCCAC ATAGATGCGT GAAGTATCTG GGACCACTGT GTCAACGAAG AAGTTACAGC 300
CCAAATCCAC CTGCATATAT AACTCCGAGT GCTTAGCTTC CTGGAGTCGC TCAATGACAT 360
10 TTCTCAGTTG AGGGTATTTG GCCAGCTGTT CATATACCTG GTCTCGATGG TCCAGAACTT 420
TCGGAAGTCC CGCTGCAGAA CGTCACTGAT GAAGGGCTCG TGGGGAGAAT TTCT 474

(2) INFORMATION FOR SEQ ID NO:63:

15

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 454 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

TGGCATCTGA AATCTTTTAT TGGAAGATCA TTGTTGTTTG CCAATTAGAA GACACAGACA 60
30 GCAGACGAAC AGTGAAAACA GAGCCCAGTG ACGAGAGCCG GCCCCTTGGT TGGGGACCCT 120
CCCCAACTAC CTGGTAGACC AGCCTGGTGA CCTCTGCCCT TCCCCGGACC CCCGGGCCTT 180
TGGCATAATG CTGATGGGGG GCTGCAGGCA GTGAAGCCCC TTGACTCAAA GCAGAGACTT 240
35 GATTGGGCGC TGGAGAGTGG AGACAGTGA GAGGCCAGGG AGGGCTGGGC GGGCCCCCA 300
GGCTGGGCGC AGCAGCGCAA GTAGAGGAAG TCAGGAGCGG GCGAGATGGC ATCTATCTTG 360
40 TTTTCTTGAA AAGGGGGCAC ATAGGGGGCC TGGGAAGCAG GTGGCGGGTG GGTAGCTTGG 420
GGAAGGTCAA CACACTGAAC ATCCTTCTTC ATCG 454

(2) INFORMATION FOR SEQ ID NO:64:

45

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
50 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

AGTGATTATG CTTTATTTA TTTCCAATT CTTATGGGTA ACATAATTTT CAGACAATGT 60

TAGCTGTTTT TAATCCATCA GTAAACTGCA TTAAGATTCT TAATAAACAA AACTGANGG 120
CCTCTTCCAT ATTGGTTTCA TCTGCATTTT TTTTATATG CTGGTCATGT GGCTTTACTT 180
5 TCAGCCTCAC TCTTTTCTTC TTCCAAATGG ATTATCCTTA AACCTTTTAC CTTTAAAGAG 240
CCTGAGATTT ATATTAACT CGAACACAG TTGGGCTCTG TTGGCCCTGT GTTCATGTTT 300
10 TCCTAAG 307

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 319 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

25 CCCCCTTTAA GTGTTACACT TTTTITAAA ACTTAACATT TCAGGAGGTC ATACGCATAC 60
ACCTCAAAC GCAAAAAATT CCAGGCATAA AACTATTAT CTGGGTTAGT TGCCCATCTT 120
30 TCTTCTCAA ATGTCAAAC GTCCACAAA AAAGTCTTAA GAAAGTCAAT TCCACTGTCC 180
ATTGGTGTGG GGTAGAAAC CTATGTCTCA TCCACTGCAT GGAATCCATG TTAAAGAAG 240
CCTGCCTTGG TTGTTTATCA TCACAGGACT CTTGTGTTAA TCCATTCTCC CTCAATTCCC 300
35 CACAGTAGAC TGCCATCTT 319

(2) INFORMATION FOR SEQ ID NO:66:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: cDNA

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GAATTCTGCG GCCGCTCCT GAGCAAAAGC CCATCCTCAC TCAGCGCTAA CATCATCAGC 60
AGCCCGAAAG GTTCTCCTTC TTCATCAAGA AAAAGTGGAA CCAGCTGTCC CTCCAGCAAA 120
55 AACAGCAGCC CTAATAGCAG CCCACGGACT TTGGGGAGGA GCAAAGGGAG GCTCCGGCTG 180
CCCCAGATTG GCAGCAAAAA TAACTGTCA AGTAGTAAAG AGAACTTGGG TGCCAGCAAA 240

5 GAAAATGGGG CTGGGCAGAT ATGTGAGCTG GCTGACGCCT TGAGTCGAGG GCATGTGCTG 300
GGGGGCAGCC AACCAGAGTT GGGTCACTCC TCAGGACCAT GAGGTAGCTT TGGGCCAATG 360
GATTCCTTTA TGAGCATGAG GAATGTAGCA ATGGTTACAG CAATGGTCAG CTTGGAACCA 420
CAGTGAGGAG AAAGCACTGA TGACCAAGAG GAGATCTTCG TTTAAGCCTA TTTATATCTA 480
10 TATGAATTCG GGCAATCAGA TTCT 504

(2) INFORMATION FOR SEQ ID NO:67:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 504 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
20 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GAATTCTGCG GCCGCCTCCT GAGCAAAAGC CCATCCTCAC TCAGCGCTAA CATCATCAGC 60
30 AGCCCGAAAG GTTCTCCTTC TTCATCAAGA AAAAGTGGAA CCAGCTGTCC CTCCAGCAAA 120
AACAGCAGCC CTAATAGCAG CCCACGGACT TTGGGGAGGA GCAAAGGGAG GCTCCGGCTG 180
CCCCAGATTG GCAGCAAAAA TAACTGTCA AGTAGTAAAG AGAACTTGA TGCCAGCAAA 240
35 GAAAATGGGG CTGGGCAGAT ATGTGAGCTG GCTGACGCCT TGAGTCGAGG GCATGTGCTG 300
GGGGGCAGCC AACCAGAGTT GGGTCACTCC TCAGGACCAT GAGGTAGCTT TGGGCCAATG 360
40 GATTCCTTTA TGAGCATGAG GAATGTAGCA ATGGTTACAG CAATGGTCAG CTTGGAACCA 420
CAGTGAGGAG AAAGCACTGA TGACCAAGAG GAGATCTTCG TTTAAGCCTA TTTATATCTA 480
TATGAATTCG GGCAATCAGA TTCT 504

45

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 365 base pairs
50 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

AACTATTTTA ATTAGAATTT TTATTTGGTG CTTCAGGGCC ACAGGATAAA ATAACTACAT 60
5 TTAGCTTGCC TTTCAGTGAC GCTTTGGCCA AATGTCAGCT ACAAGGAGTC ATCTCCCTCA 120
CCGCCAAGCT GTCTAGCAGC CAGAGTGGTA GCTTTACTGT AACACACAGT ACTTTTGGTA 180
ATCAGACTCA AAGTCTTCAT CCATACTGCT TGTGTCTGCC ATCTTTTGGG CATCAGTCTT 240
10 GGGCAGAAAT TGTGCATAGT CTATCCCCTG CTGCTCATAG AAAAGATTGT AGGCAGAGTC 300
GGGTGTCAAT TTCATCCGGG TGAAGTTCCT TACAGCTGCT GTCATTGTAC AAGTACCACT 360
15 TGCAG 365

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 444 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

30 GAATTCTGCG GCCGNCGGGC ACAGGCAGTG CTGGAGGAAG ACCACTACGG GATGGAGGAC 60
GTCAGGAAAC GCATCCTGGA GTTCATNGCC GTTAGCCAGC TCCGCGGNTC CACCCAGGGC 120
35 AAGATCCTCT GCTTCTATGG CCCCCCTGGC GTGGGTAAGA CCAGCATTGG TCGCTCCATC 180
GNCCGCGCCT GACCGAGAGT ACTTCCCCTG TCAGNGTCGG GGGGATTATG ACGTNGGTGA 240
GATCAAAGGG CACAGGGGGC CTCCGTGGGC GCCATTCCGG AAGATCATCC ANTNTTGGGG 300
40 AAGACCAAAN GGNGAACCCC TTATTCNCA TCGAGAAGGN GGNAAAAATC GNCCANGTTA 360
CNAGGGGGCC CCNNNTCGNA ATTNTTNTGT TTTTITACCA ANAAAAATNT CATTTCCCNG 420
45 ACCNTNCTGG GGTCCCCTN ANTT 444

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 423 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

ACTGAAAATG ACTTTAATCA TTAAATAGCT TCTATGCCAC ACTCTGATTA AGCCGACTGA 60
5 GGTCCCTGGG ATCTGGGTCA CTGGACCGAG CTGCTCGCTC GGTGGCTCCA CTGCCAGGTC 120
CGGGCGCGCT CCCACAGGG GTCAGTCTTG GCCAGACAGG GCTGANATCC GCGCCTGAAG 180
10 TCCGGGTGGG CCGCACCGTC CACGGCAGGG CTCTGCTTTC GCCGGGAGGG GAAGTCGAGG 240
TCTCCCGNNG GGTCCAGAAG GGGAACCCCA GGCCCCGGGG ATNAANGTNC CAGGCGGGAA 300
AGTCCCCCTT TCTCNGTTGG AANAAAAAAA AANAACCCCN NGNGCTTGGG NNAAAGGCCT 360
15 NCTCCTGGNG GNCNACANAN NAAGATNTTN CCCGNGGGGG ATTCCCCAAA NAAANCAAAT 420
TTT 423

(2) INFORMATION FOR SEQ ID NO:71:

20

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
25 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TACCAGCCTC TTGCTGAGTG GAGA 24

35 (2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
40 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TAGACAAGCC GACAACCTTG ATTG 24

1. A substantially pure preparation of a CDK4-binding protein, or a fragment thereof, comprising an amino acid sequence at least 60% homologous to a polypeptide selected from a group consisting of SEQ ID Nos. 25-48.
- 5 2. A preparation of a purified or recombinant polypeptide comprising an amino acid sequence identical or homologous to a sequence of SEQ ID No. 31, which polypeptide binds to a cyclin dependent kinase.
3. The preparation of claim 2, which polypeptide functions in one of either role of an agonist or an antagonist of cell cycle regulation by a cyclin-dependent kinase (CDK).
10
4. The preparation of claim 2, which polypeptide has a proteolytic activity.
5. The preparation of claim 4, which polypeptide binds CDK4.
15
6. The preparation of claim 4, which polypeptide is a fusion protein.
7. A preparation of a purified or recombinant polypeptide comprising an amino acid sequence identical or homologous to a sequence of SEQ ID No. 33, which polypeptide binds to a cyclin dependent kinase.
20
8. The preparation of claim 7, which polypeptide functions in one of either role of an agonist or an antagonist of cell cycle regulation by cyclin-dependent kinase (CDK).
- 25 9. The preparation of claim 7, which polypeptide has an isopeptidase activity.
10. The preparation of claim 9, which polypeptide is a de-ubiquitinating enzyme.
11. The preparation of claim 7, which polypeptide is a fusion protein.
30
12. A preparation of a purified or recombinant polypeptide comprising an amino acid sequence identical or homologous to a sequence of SEQ ID No. 43, which polypeptide binds to a cyclin dependent kinase.
- 35 13. The preparation of claim 12, which polypeptide functions in one of either role of an agonist or an antagonist of cell cycle regulation by a cyclin-dependent kinase (CDK).
14. The preparation of claim 12, which polypeptide has a kinase activity.

15. The preparation of claim 14, which polypeptide is a stress-activated protein kinase.
16. The preparation of claim 12, which polypeptide is a fusion protein.
- 5 17. A preparation of a purified or recombinant polypeptide comprising an amino acid sequence identical or homologous to a sequence of SEQ ID No. 45, which polypeptide binds to a cyclin dependent kinase.
- 10 18. The preparation of claim 17, which polypeptide functions in one of either role of an agonist or an antagonist of cell cycle regulation by a cyclin-dependent kinase (CDK).
19. The preparation of claim 17, which polypeptide is cdc37 homolog.
- 15 20. The preparation of claim 17, which polypeptide binds CDK4.
21. The preparation of claim 17, which polypeptide is a fusion protein.
22. An antibody preparation specifically reactive with an epitope of the polypeptide of
20 claim 1.
23. An antibody preparation specifically reactive with an epitope of the polypeptide of claim 2.
- 25 24. An antibody preparation specifically reactive with an epitope of the polypeptide of claim 7.
25. An antibody preparation specifically reactive with an epitope of the polypeptide of claim 12.
- 30 26. An antibody preparation specifically reactive with an epitope of the polypeptide of claim 17.
27. A polypeptide a recombinantly produced from a pJG4-5-CDKBP clone of ATCC
35 deposit no. 75788.

28. An nucleic acid having a nucleotide sequence which encodes a polypeptide comprising an amino acid sequence identical or homologous to a sequence of one of SEQ ID No. 25-47, which polypeptide binds to a cyclin dependent kinase.
- 5 29. The nucleic acid of claim 28, wherein said polypeptide encoded by said nucleic acid functions in one of either role of an agonist of cell cycle regulation or an antagonist of cell cycle regulation.
- 10 30. The nucleic acid of claim 28, wherein said nucleotide sequence hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides of one of SEQ ID Nos. 1-24 and 49-70.
- 15 31. The nucleic acid of claim 28, wherein said polypeptide comprises an amino acid sequence identical or homologous to a sequence of SEQ ID No. 31.
32. The nucleic acid of claim 28, wherein said polypeptide comprises an amino acid sequence identical or homologous to a sequence of SEQ ID No. 33.
- 20 33. The nucleic acid of claim 28, wherein said polypeptide comprises an amino acid sequence identical or homologous to a sequence of SEQ ID No. 43.
34. The nucleic acid of claim 28, wherein said polypeptide comprises an amino acid sequence identical or homologous to a sequence of SEQ ID No. 45.
- 25 35. The nucleic acid of claim 28, wherein said polypeptide is a fusion protein.
36. The nucleic acid of claim 28, further comprising a transcriptional regulatory sequence operably linked to said nucleotide sequence so as to render said nucleotide sequence suitable for use as an expression vector.
- 30 37. An expression vector, capable of replicating in at least one of a prokaryotic cell and eukaryotic cell, comprising the nucleic acid of claim 36.
- 35 38. A host cell transfected with the expression vector of claim 37 and expressing said polypeptide.

39. A method of producing a recombinant CDK4-binding protein comprising culturing the cell of claim 38 in a cell culture medium to express said CDK4-binding protein and isolating said CDK4-binding protein from said cell culture.
- 5 40. A transgenic animal comprising cells harboring a recombinant form the nucleic acid of claim 28.
-
41. The nucleic acid of claim 28, which includes intronic nucleotide sequences disrupting said polypeptide-encoding sequence.
- 10 42. A nucleic acid composition comprising, as nucleic acid component, a substantially purified oligonucleotide, said oligonucleotide containing a region of nucleotide sequence which hybridizes under stringent conditions to at least 40 consecutive nucleotides of sense or antisense sequence selected from a group consisting of SEQ ID Nos. 1-24 and 49-70, or naturally occurring mutants thereof.
- 15 43. The nucleic acid composition of claim 42, which oligonucleotide hybridizes under stringent conditions to at least 80 consecutive nucleotides of sense or antisense sequences selected from a group consisting of SEQ ID Nos. 1-24 and 49-70, or naturally occurring mutants thereof.
- 20 44. The nucleic acid composition of claim 42, which oligonucleotide further comprises a label group attached thereto and able to be detected.
- 25 45. The nucleic acid composition of claim 42, which oligonucleotide has at least one non-hydrolyzable bond between two adjacent nucleotide subunits.
46. A diagnostic test kit for identifying an transformed cells, comprising the nucleic acid of claim 42, for measuring a level of a nucleic acid encoding a CDK-binding protein in a sample of cells isolated from a patient.
- 30 47. An assay for screening test compounds for an inhibitor of an interaction of a cyclin dependent kinase (CDK) with a CDK4-binding protein (CDK-BP) comprising
- 35 i. combining a CDK and a CDK4-binding protein, which CDK4-binding protein includes an amino acid sequence represented in a group consisting of SEQ ID Nos. 25-48, under conditions wherein said CDK and said CDK4-binding protein are able to interact;
- ii. contacting said combination with a test compound; and

SUBSTITUTE SHEET (RULE 26)

iii. detecting the formation of a complex comprising said CDK and said CDK4-binding protein,
wherein a statistically significant decrease in the formation of said complex in the presence of said test compound is indicative of an inhibitor of the interaction between
5 said CDK and said CDK4-binding protein.

48. A method of identifying an agent which disrupts the ability of a CDK4-binding protein to regulate a eukaryotic cell cycle, comprising:
- 10 i. providing an interaction trap assay system including a first fusion protein comprising a cyclin-dependent kinase (CDK) and second fusion protein comprising a CDK4-binding protein including an amino acid sequence selected from a group consisting of SEQ ID Nos. 25-48, under conditions wherein said interaction trap assay is sensitive to interactions between the CDK of said first fusion protein and said CDK4-binding protein of said
15 second polypeptide;
 - ii. contacting said interaction trap assay with a candidate agent;
 - iii. measuring a level of interactions between said fusion proteins in the presence of said candidate agent; and
 - 20 iv. comparing the level of interaction of said fusion proteins in the presence of said candidate agent to a level of interaction of said fusion proteins in the absence of the candidate agent,

wherein a decrease in the level of interaction in the presence of said candidate agent is indicative of inhibition of an interaction between said CDK and said CDK-binding protein.

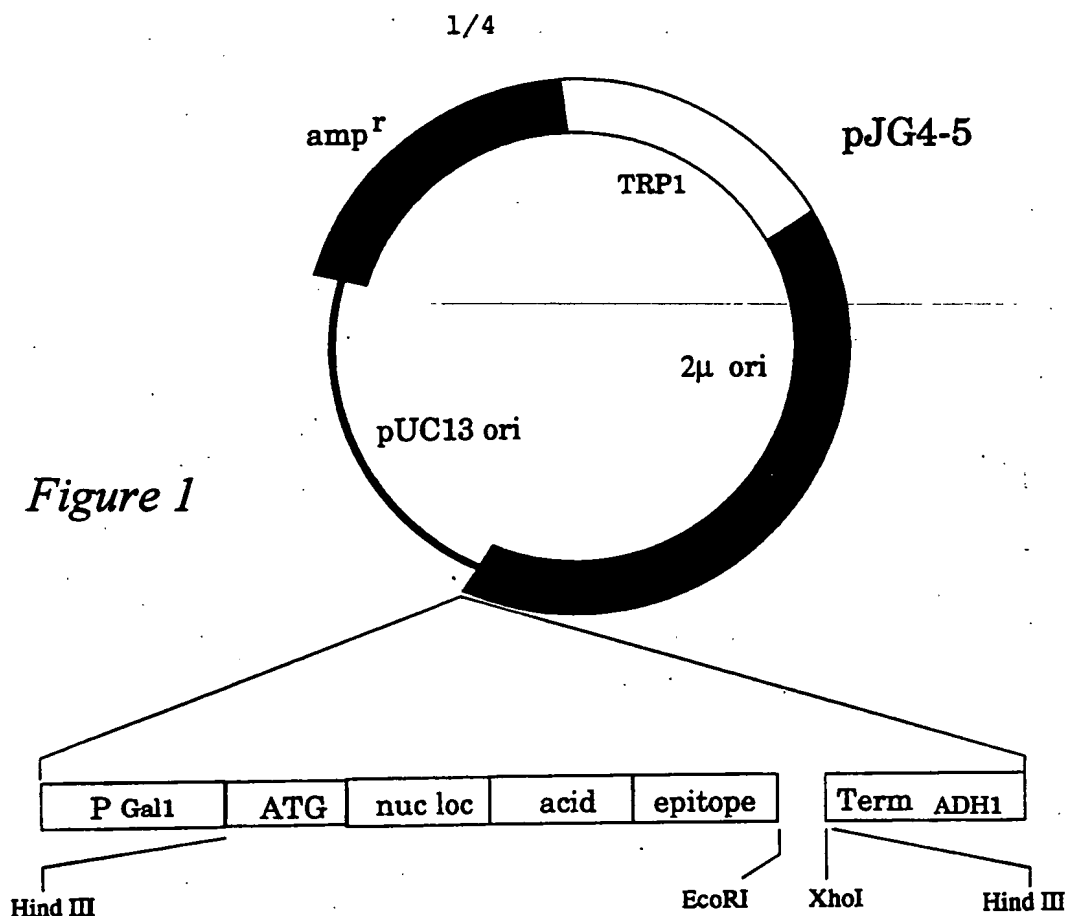
- 25 49. A method of determining if a subject is at risk for a disorder characterized by unwanted cell proliferation, comprising detecting, in a tissue of said subject, the presence or absence of a genetic lesion characterized by at least one of
a mutation of a gene encoding a protein selected from a group consisting of
30 SEQ ID Nos. 25-48, or homologs thereof; and the mis-expression of said gene.

50. The method of claim 49, wherein detecting said genetic lesion comprises ascertaining the existence of at least one of
- 35 i. a deletion of one or more nucleotides from said gene,
 - ii. an addition of one or more nucleotides to said gene,
 - iii. an substitution of one or more nucleotides of said gene,
 - iv. a gross chromosomal rearrangement of said gene.
 - v. a gross alteration in the level of a messenger RNA transcript of said gene,

SUBSTITUTE SHEET (RULE 26)

- vi. the presence of a non-wild type splicing pattern of a messenger RNA transcript of said gene, and
- vii. a non-wild type level of said protein.

- 5 51. The method of claim 49, wherein detecting said genetic lesion comprises
- 10 i. providing a probe/primer comprising an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of nucleic acid of one of SEQ ID Nos. 1-24 and 49-70, or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with said gene;
 - ii. exposing said probe/primer to nucleic acid of said tissue; and
 - iii. detecting, by hybridization of said probe/primer to said nucleic acid, the presence or absence of said genetic lesion.
- 15 52. The method of claim 49, wherein detecting said lesion comprises utilizing said probe/primer to determine the nucleotide sequence of said gene and, optionally, of said flanking nucleic acid sequences.
- 20 53. The method of claim 49, wherein detecting said lesion comprises utilizing said probe/primer to in a polymerase chain reaction (PCR) or ligation chain reaction (LCR).
- 25 54. The method of claim 50, wherein the level of said protein is detected in an immunoassay.



AAG CTT ATG GGT GCT CCT CCA AAA AAG AAG AGA AAG GTA GCT GGT
M G A P P K K K R K V A G

ATC AAT AAA GAT ATC GAG GAG TGC AAT GCC ATC ATT GAG CAG TTT
I N K D I E E C N A I I E Q F

ATC GACTAC CTG CGC ACC GGA CAG GAG ATG CCG ATG GAA ATG GCG
I D Y L R T G Q E M P M E M A

GAT CAG GCGATT AAC GTG GTG CCG GGC ATG ACG CCG AAA ACC ATT
D Q A I N V V P G M T P K T I

CTT CAC GCC GGGCCG CCG ATC CAG CCT GAC TGG CTG AAA TCG AAT
L H A G P P I Q P D W L K S N

GGT TTT CAT GAA ATTGAA GCG GAT GTT AAC GAT ACC AGC CTC TTG
G F H E I E A D V N D T S L L

CTG AGT GGA GAT GCC TCCTAC CCT TAT GAT GTG CCA GAT TAT GCC
L S G D A S Y P Y D V P D Y A

SUBSTITUTE SHEET (RULE 26)

Figure 2

pjG4-5 clone	CDK2	CDK3	CDK4	CDK5	CDK6	cdi1	Rb	Δ Rb	p53	cycC	cycD1	cycD2	cycE
#11			++	+	++								
#13			++	++	++								
#22			++	+			+	+	++				
#36			++				+		+				
#61			++	+	+								
#68			++	+	++								
#71			++	+	+								
#75			++										
#116			++		+		++	++	++				
#118			++										
#121			++										
#125			++			+	+	+	+				
#127			++				++	++	++				
#165			++				+	+	+				
#166			+		+		++	+	++				
#190			+				+	+	++				
#193			+		+		+	+	++				
#216			+		+		+/-	+/-	+/-				
#225			+		+	+							
#227			+		+								
#267			+		+								
#269			+		+/-		++	++	++				
#295			+		+		+/-	+/-					
#297			+		+		+/-	+/-	+/-				
Cyclin D1	++	++	++	+	++								
Cyclin D3	++	++	++		++								
p16			++		++								
p21	++	++	+		+						+	+	

SUBSTITUTE SHEET (RULE 26)

3/4

Figure 3

cDNA	spln	thy	pros	tes	ov	s.i.	co	p.b.	he	br	pla	lu	lvr	s.m.	ki	pa	mRNA size (kb)
cyclin B1	+/-	+++	-	+++	+	+	+	+/-	+/-	+	+	+/-	-	-	+	-	
actin	+++	+++	+++	+++	+++	+++	+++	+++	+	+	+	+++	+++	+++	+++	+	
#11	+	+	-	-	+	+	+	-	-	-	+++	+	-	+	+	-	1.2
#13	+++	+++	+++	+++	+++	+++	+++	+++	+	+	+	+	+	+	+	+	1.3
	+/-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	3.3
	-	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	5.0
#22	++	++	+	++	+	+	+	++	+	+	++	+	+	+	+	+	8.4
#36	+++	+++	+++	+	+++	+++	+++	+++	+++	+	++	+++	+++	+++	+++	++	9.5
	+/-	+/-	+/-	+/-	+/-	+/-	+/-	-	+	+/-	++	+	+	++	++	++	1.3
#61	-	-	-	-	-	+/-	+/-	-	+	+	++	+	-	+	+	+	9.5
#68	+	++	-	+	-	+/-	+/-	++	+	++	++	+	-	+	+	+	5.5
#71	+	+	+	+	-	-	-	++	+	++	-	+	-	+	+/-	+	1.2
#75	++	++	+	++	++	+	++	++	++	+	+	++	++	++	++	+	3.0
#118	++	+++	++	++	++	++	++	++	+	+	+	+	+	+	+	+	2.1
#121	+	+	++	-	+	++	++	-	-	++	-	-	++	-	++	-	5.5
	+	+	++	+	+++	+++	+++	-	++	-	+	+	-	+	+	+	2.4
#125	-	-	+	-	+++	+++	+++	-	++	-	-	-	-	-	-	-	3.7
#127	-	-	-	+/-	-	-	++	-	+	-	++	+	-	+	+	-	5.0
#165	+/-	+	+	+	+	+/-	+/-	-	++	-	++	+	+	+	+	=	3.9
#166	++	++	++	++	++	++	++	++	+++	++	++	+++	+	+++	+++	+++	1.1
#216	+	++	+	++	++	+	+	+	++	++	++	+	+	++	++	++	1.3
	++	++	+	++	++	+	+	+	+	++	++	+	+	++	++	++	8.0
#225	+/-	+/-	++	++	++	++	++	++	+	++	+	+	+	++	+	++	4.0
	++	++	++	++	++	++	++	++	+/-	+	+/-	-	-	+	+/-	+	3.0
	++	++	++	++	++	++	++	++	++	++	++	++	+	++	++	++	2.0

[illegible]

sp ln=spleen; thy=thymus; pros=prostate; tes=testis; ov=ovary; s.i.=small intestine; co=colon; p.b.=peripheral blood; he=heart; br=brain; pla=placenta; lu=lung; lvr=liver; s.m.=smooth muscle; ki=kidney; pa=pancreas

PCT

WORLD INTELLECTUAL PROP
International I

INTERNATIONAL APPLICATION PUBLISHED UNDER

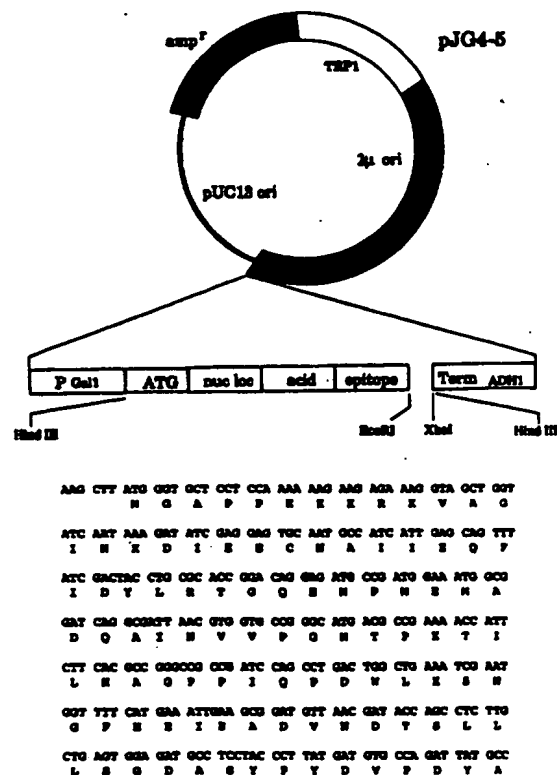
WO 9533819A3

(51) International Patent Classification ⁶ : C12N 15/12, C07K 14/47, 16/18, A01K 67/027, G01N 33/68, C12Q 1/68		A3	(11) International Publication Number: WO 95/33819
			(43) International Publication Date: 14 December 1995 (14.12.95)
(21) International Application Number: PCT/US95/07113		(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 2 June 1995 (02.06.95)			
(30) Priority Data: 08/253,155 2 June 1994 (02.06.94) US		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant: MITOTIX, INC. [US/US]; Building 600, One Kendall Square, Cambridge, MA 02139 (US).		(88) Date of publication of the international search report: 21 March 1996 (21.03.96)	
(72) Inventors: DRAETTA, Giulio; 27 Arlington Street, Winchester, MA 01890 (US). GYURIS, Jeno; 11 Cardinal Street, Winchester, MA 01890 (US).			
(74) Agents: VINCENT, Matthew, P. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).			

(54) Title: CDK4 (CYCLIN DEPENDENT KINASE 4) BINDING PROTEINS

(57) Abstract

The present invention relates to the discovery of novel proteins of mammalian origin which can associate with the human cyclin dependent kinase 4 (CDK4).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

INTERNATIONAL SEARCH REPORT

Int. Application No
PCT/US 95/07113

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C07K16/18 A01K67/027 G01N33/68
C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	J BIOL CHEM, MAY 6 1994, 269 (18) P13279-88, UNITED STATES, AZZI L ET AL 'Purification of a 15-kDa cdk4- and cdk5-binding protein.' see the whole document ---	1-22, 27-30, 35-54
A	CELL, vol. 71, 30 October 1992 NA US, pages 505-514, XIONG, Y 'D type cyclins associate with multiple protein kinases and the DNA replication and repair factor PCNA' cited in the application see page 509, column 1, line 12 - page 509, column 1, line 55 -----	1,22, 27-30, 35-54

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *B* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *a* document member of the same patent family

Date of the actual completion of the international search

13 February 1996

Date of mailing of the international search report

20. 02. 96

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+ 31-70) 340-3016

Authorized officer

Nauche, S

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 95/07113

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

- 20 subjects

See additional sheets PCT/ISA/210

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1,22,27-30,35,54 partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US95/07113

FURTHER INFORMATION CONTINUED FROM PCT/SAJ210

- | | | |
|---|--|--|
| 1 | 1.22. 27-30.35-54
partially | CDK4-binding protein having a selectin-like activity, fragments thereof, comprising an amino acid at least 60% homologous to a selectin like CDK4-binding protein as described in sequence 25, corresponding nucleic acids, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 2 | 7-11.24.32 totally
1.22. 27-30.35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequences SEQ IDs No. 26,27,44 [substrate and/or inhibitor of the CDK4 kinase activity], corresponding nucleic acids, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 3 | 1.22. 27-30.35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequences SEQ IDs No. 29, 40 [cytoskeletal elements] corresponding nucleic acids, mutants thereof and fusion proteins.
Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals |
| 4 | 2-6.23.31 totally
1.22. 27-30.35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequence SEQ ID No. 31[protease], corresponding nucleic acids, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals |
| 5 | 7-11.24.32 totally
1.22. 27-30.35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequence SEQ ID No. 33[isopeptidase], corresponding nucleic acids, mutants thereof and fusion proteins.
Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 6 | 1.22. 27-30.35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequences SEQ IDs No. 34,37 [DNA binding proteins], corresponding nucleic acids, mutants thereof and fusion proteins.
Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 7 | 1.22. 27-30.35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequence SEQ ID No. 42 [mRNA splicing element], corresponding nucleic acids, mutants thereof and fusion proteins.
Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US95/07113

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

- | | | |
|----|---|---|
| 8 | 12-16,25,33 totally
1,22, 27-30,35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequence SEQ ID No. 43[kinase], corresponding nucleic acids, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals |
| 9 | 17-21,26,34 totally
1,22, 27-30,35-54
partially | Polypeptide comprising an amino acid identical or at least 60% homologous to sequence SEQ ID No. 45[cdc37 homolog], corresponding nucleic acids, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 10 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide selected from the group consisting of SEQ ID 28, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 11 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 30, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 12 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 32 corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 13 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 35, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 14 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 36 corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US95/07113

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

- | | | |
|----|--------------------------------|--|
| 15 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 38, corresponding nucleic acids, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 16 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 39 corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 17 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 41, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 18 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 461, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 19 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 47, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |
| 20 | 1,22, 27-30,35-54
partially | CDK4-binding protein, fragments thereof, comprising an amino acid at least 60% to a polypeptide consisting of SEQ ID 48, corresponding nucleic acid, mutants thereof and fusion proteins. Antibodies to the polypeptides
Screening tests, assays involving said defined sequences. Non-human transgenic animals. |